which in this case is the vector $(p_1\zeta_1, \dots, p_m\zeta_m)$, is not the zero vector, then the optimal solution must satisfy the Lagrange multiplier equations (3.3.22). This gradient is not the zero vector, so this hypothesis is satisfied. However, even when this hypothesis is satisfied, the theorem does not guarantee that there is an optimal solution; the solution to the Lagrange multiplier equations may in fact minimize the expected utility. The solution could also be neither a maximizer nor a minimizer. Therefore, in this exercise, we outline a different method for verifying that the random variable X_N given by (3.3.25) maximizes the expected utility.

We begin by changing the notation, calling the random variable given by (3.3.25) X_N^* rather than X_N. In other words,

$$X_N^* = I\left(\frac{\lambda}{(1+r)^N} Z\right), \tag{3.6.1}$$

where λ is the solution of equation (3.3.26). This permits us to use the notation X_N for an arbitrary (not necessarily optimal) random variable satisfying (3.3.19). We must show that

$$\mathbb{E}U(X_N) \le \mathbb{E}U(X_N^*). \tag{3.6.2}$$

(i) Fix $y > 0$, and show that the function of x given by $U(x) - yx$ is maximized by $y = I(x)$. Conclude that

$$U(x) - yx \le U(I(y)) - yI(y) \text{ for every } x. \tag{3.6.3}$$

(ii) In (3.6.3), replace the dummy variable x by the random variable X_N and replace the dummy variable y by the random variable $\frac{\lambda Z}{(1+r)^N}$. Take expectations of both sides and use (3.3.19) and (3.3.26) to conclude that (3.6.2) holds.

Exercise 3.9 (Maximizing probability of reaching a goal). (Kulldorf [30], Heath [19])

A wealthy investor provides a small amount of money X_0 for you to use to prove the effectiveness of your investment scheme over the next N periods. You are permitted to invest in the N-period binomial model, subject to the condition that the value of your portfolio is never allowed to be negative. If at time N the value of your portfolio X_N is at least γ, a positive constant specified by the investor, then you will be given a large amount of money to manage for her. Therefore, your problem is the following:

Maximize

$$\mathbb{P}(X_N \ge \gamma),$$

where X_N is generated by a portfolio process beginning with the initial wealth X_0 and where the value X_n of your portfolio satisfies

$$X_n \ge 0, \; n = 1, 2, \dots, N.$$

In the way that Problem 3.3.1 was reformulated as Problem 3.3.3, this problem may be reformulated as

Maximize

$$\mathbb{P}(X_N \ge \gamma)$$

subject to

$$\widetilde{\mathbb{E}}\frac{X_N}{(1+r)^N} = X_0,$$
$$X_n \ge 0, \; n = 1, 2, \dots, N.$$

(i) Show that if $X_N \ge 0$, then $X_n \ge 0$ for all n.

(ii) Consider the function

$$U(x) = \begin{cases} 0, \text{ if } 0 \le x < \gamma, \\ 1, \text{ if } x \ge \gamma. \end{cases}$$

Show that for each fixed $y > 0$, we have

$$U(x) - yx \le U(I(y)) - yI(y) \;\forall x \ge 0,$$

where

$$I(y) = \begin{cases} \gamma, \text{ if } 0 < y \le \frac{1}{\gamma}, \\ 0, \text{ if } y > \frac{1}{\gamma}. \end{cases}$$

(iii) Assume there is a solution λ to the equation

$$\mathbb{E}\left[\frac{Z}{(1+r)^N} I\left(\frac{\lambda Z}{(1+r)^N}\right)\right] = X_0. \tag{3.6.4}$$

Following the argument of Exercise 3.8, show that the optimal X_N is given by

$$X_N^* = I\left(\frac{\lambda Z}{(1+r)^N}\right).$$

(iv) As we did to obtain Problem 3.3.5, let us list the $M = 2^N$ possible coin toss sequences, labeling them $\omega^1, \dots, \omega^M$, and then define $\zeta_m = \zeta(\omega^m)$, $p_m = \mathbb{P}(\omega^m)$. However, here we list these sequences in ascending order of ζ_m i.e., we label the coin toss sequences so that

$$\zeta_1 \le \zeta_2 \le \cdots \le \zeta_M.$$

Show that the assumption that there is a solution λ to (3.6.4) is equivalent to assuming that for some positive integer K we have $\zeta_K < \zeta_{K+1}$ and

$$\sum_{m=1}^{K} \zeta_m p_m = \frac{X_0}{\gamma}. \tag{3.6.5}$$

(v) Show that X_N^* is given by

$$X_N(\omega^m) = \begin{cases} \gamma, \text{ if } m \le K, \\ 0, \text{ if } m \ge K+1. \end{cases}$$

4

American Derivative Securities

4.1 Introduction

European option contracts specify an expiration date, and if the option is to be exercised at all, the exercise must occur on the expiration date. An option whose owner can choose to exercise at any time up to and including the expiration date is called *American*. Because of this early exercise feature, such an option is at least as valuable as its European counterpart. Sometimes the difference in value is negligible or even zero, and then American and European options are close or exact substitutes. We shall see in this chapter that the early exercise feature for a call on a stock paying no dividend is worthless; American and European calls have the same price. In other cases, most notably put options, the value of this early exercise feature, the so-called *early exercise premium*, can be substantial. An intermediate option between American and European is *Bermudan*, an option that permits early exercise but only on a contractually specified finite set of dates.

Because an American option can be exercised at any time prior to its expiration, it can never be worth less than the payoff associated with immediate exercise. This is called the *intrinsic value* of the option.

In contrast to the case for a European option, whose discounted price process is a martingale under the risk-neutral measure, the discounted price process of an American option is a supermartingale under this measure. The holder of this option may fail to exercise at the optimal exercise date, and in this case the option has a tendency to lose value; hence, the supermartingale property. During any period of time in which it is not optimal to exercise, however, the discounted price process behaves like a martingale.

To price an American option, just as with a European option, we shall imagine selling the option in exchange for some initial capital and then consider how to use this capital to hedge the short position in the option. In this case, we need to be ready to pay off the option at all times prior to the expiration date because we do not know when it will be exercised. We determine when, from our point of view, is the worst time for the owner to exercise the option. From the owner's point of view, this is the *optimal exercise time*, and we shall call it that. We then compute the initial capital we need in order to be hedged against exercise at the optimal exercise time. Finally, we show how to invest this capital so that we are hedged even if the owner exercises at a nonoptimal time. We conclude that the initial price of the option is the capital required to be hedged against optimal exercise.

Section 4.2 sets out the basic American derivative security pricing algorithm when the payoff of the derivative is not path-dependent (i.e., the payoff depends only on the current value of the underlying asset, not on any previous values). In order to develop a complete theory of American derivative securities that includes path-dependent securities, we need the notion of *stopping time*, which is introduced in Section 4.3. Armed with this concept, we work out the general theory of American derivative securities in Section 4.4. One consequence of this general theory is that there is no gain from early exercise of an American call on a nondividend-paying stock. This and related results are developed in Section 4.5.

4.2 Non-Path-Dependent American Derivatives

In this section, we develop a pricing algorithm for American derivative securities when the payoff is not path dependent. We first review the pricing algorithm for European derivative securities when the payoff is not path dependent. In an N-period binomial model with up factor u, down factor d, and interest rate r satisfying the no-arbitrage condition $0 < d < 1 + r < u$, consider a derivative security that pays off $g(S_N)$ at time N for some function g. Because the stock price is Markov, we can write the value V_n of this derivative security at each time n as a function v_n of the stock price at that time i.e., $V_n = v_n(S_n)$, $n = 0, 1, \ldots, N$ (Theorem 2.5.8 of Chapter 2). The risk-neutral pricing formula (see (2.4.12) and its Markov simplification (2.5.2) of Chapter 2) implies that, for $0 \leq n \leq N$, the function v_n is defined by the *European algorithm*:

$$v_N(s) = \max\{g(s), 0\}, \tag{4.2.1}$$

$$v_n(s) = \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us) + \tilde{q}v_{n+1}(ds)\Big], \quad n = N-1, N-2, \ldots, 0, \tag{4.2.2}$$

where $\tilde{p} = \frac{1+r-d}{u-d}$ and $\tilde{q} = \frac{u-1-r}{u-d}$ are the risk-neutral probabilities that the stock goes up and down, respectively. The replicating portfolio (which hedges a short position in the option) is given by (see (1.2.17) of Chapter 1)

$$\Delta_n = \frac{v_{n+1}(uS_n) - v_{n+1}(dS_n)}{(u-d)S_n}, \quad n = 0, 1, \ldots, N. \tag{4.2.3}$$

Now consider an *American derivative security*. Again, a payoff function g is specified. In any period $n \leq N$, the holder of the derivative security can

exercise and receive payment $g(S_n)$. (In this section, the payoff depends only on the current stock price S_n at the time of exercise, not on the stock price path.) Thus, the portfolio that hedges a short position should always have value X_n satisfying

$$X_n \geq g(S_n), \quad n = 0, 1, \ldots, N. \tag{4.2.4}$$

The value of the derivative security at each time n is at least as much as the so-called *intrinsic value* $g(S_n)$, and the value of the replicating portfolio at that time must equal the value of the derivative security.

This suggests that to price an American derivative security, we should replace the European algorithm (4.2.2) by the *American algorithm*:

$$v_N(s) = \max\{g(s), 0\}, \tag{4.2.5}$$

$$v_n(s) = \max\left\{g(s), \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us) + \tilde{q}v_{n+1}(ds)\Big]\right\}, \tag{4.2.6}$$

$$n = N-1, N-2, \ldots, 0.$$

Then $V_n = v_n(S_n)$ would be the price of the derivative security at time n.

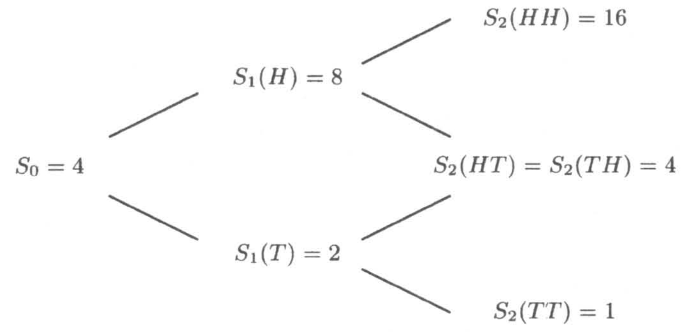

Fig. 4.2.1. A two-period model.

Example 4.2.1. In the two-period model of Figure 4.2.1, let the interest rate be $r = \frac{1}{4}$, so the risk-neutral probabilities are $\tilde{p} = \tilde{q} = \frac{1}{2}$. Consider an American put option, expiring at time two, with strike price 5. In other words, if the owner of the option exercises at time n, she receives $5 - S_n$. We take $g(s) = 5 - s$, and the American algorithm (4.2.5), (4.2.6) becomes

$$v_2(s) = \max\{5 - s, 0\},$$

$$v_n(s) = \max\left\{5 - s, \frac{2}{5}\Big[v_{n+1}(2s) + v_{n+1}\Big(\frac{s}{2}\Big)\Big]\right\}, \quad n = 1, 0.$$

In particular,

$$
\begin{aligned}
v_2(16) &= 0,\\
v_2(4) &= 1,\\
v_2(1) &= 4,\\
v_1(8) &= \max\left\{(5-8), \frac{2}{5}(0+1)\right\} = \max\{-3, 0.40\} = 0.40,\\
v_1(2) &= \max\left\{(5-2), \frac{2}{5}(1+4)\right\} = \max\{3, 2\} = 3,\\
v_0(4) &= \max\left\{(5-4), \frac{2}{5}(0.40+3)\right\} = \max\{1, 1.36\} = 1.36.
\end{aligned}
$$

This algorithm gives a different result than the European algorithm in the

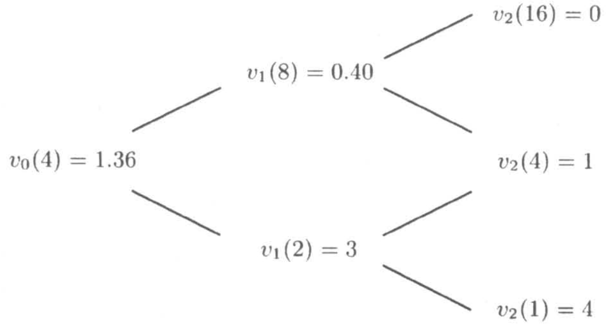

Fig. 4.2.2. American put prices.

computation of $v_1(2)$, where the discounted expectation of the time two option price, $\frac{2}{5}(1+4)$, is strictly smaller than the intrinsic value. Because $v_1(2)$ is strictly greater than the price of a comparable European put, the initial price $v_0(4)$ for the American put is also strictly greater than the initial price of a comparable European put. Figure 4.2.2 shows that American put prices.

Let us now construct the replicating portfolio. We begin with initial capital 1.36 and compute Δ_0 so that the value of the hedging portfolio at time one agrees with the option value. If the first toss results in a head, this requires that

$$
\begin{aligned}
0.40 &= v_1(S_1(H))\\
&= S_1(H)\Delta_0 + (1+r)(X_0 - \Delta_0 S_0)
\end{aligned}
$$

$$
\begin{aligned}
&= 8\Delta_0 + \frac{5}{4}(1.36 - 4\Delta_0)\\
&= 3\Delta_0 + 1.70,
\end{aligned}
$$

which implies that $\Delta_0 = -0.43$. On the other hand, if the first toss results in a tail, we must have

$$
\begin{aligned}
3 &= v_1(S_1(T))\\
&= S_1(T)\Delta_0 + (1+r)(X_0 - \Delta_0 S_0)\\
&= 2\Delta_0 + \frac{5}{4}(1.36 - 4\Delta_0)\\
&= -3\Delta_0 + 1.70,
\end{aligned}
$$

which also implies that $\Delta_0 = -0.43$. We also could have found this value of Δ_0 by substituting into (4.2.3):

$$
\Delta_0 = \frac{v_1(8) - v_1(2)}{8-2} = \frac{0.40 - 3}{8-2} = -0.43.
$$

In any case, if we begin with initial capital $X_0 = 1.36$ and take a position of Δ_0 shares of stock at time zero, then at time one we will have $X_1 = V_1 = v_1(S_1)$, regardless of the outcome of the coin toss.

Let us assume that the first coin toss results in a tail. It may be that the owner of the option exercises at time 1, in which case we deliver to her the \$3 value of our hedging portfolio and no further hedging is necessary. However, the owner may decline to exercise, in which case the option is still alive and we must continue hedging.

We consider in more detail the case where the owner does not exercise at time 1 after a first toss resulting in tail. We note that next period the option will be worth $v_2(4) = 1$ if the second toss results in head and worth $v_2(1) = 4$ if the second toss results in tail. The risk-neutral pricing formula says that to construct a hedge against these two possibilities, at time 1 we need to have a hedging portfolio valued at

$$
\frac{2}{5}\big(v_2(4) + v_2(1)\big) = 2,
$$

but we have a hedging portfolio valued at $v_1(2) = 3$. Thus, we may consume \$1 and continue the hedge with the remaining \$2 value in our portfolio. As this suggests, the option holder has let an optimal exercise time go by.

More specifically, we consume \$1 and change our position to $\Delta_1(T)$ shares of stock. If the second coin toss results in head, we want

$$
\begin{aligned}
1 &= v_2(S_2(TH))\\
&= 4\Delta_1(T) + \frac{5}{4}(2 - 2\Delta_1(T))\\
&= 1.5\Delta_1(T) + 2.50,
\end{aligned}
$$

and this implies $\Delta_1(T) = -1$. If the second coin toss results in tail, we want

$$\begin{aligned} 4 &= v_2(S_2(TT)) \\ &= \Delta_1(T) + \frac{5}{4}(2 - 2\Delta_1(T)) \\ &= -1.5\Delta_1(T) + 2.50, \end{aligned}$$

and this also implies $\Delta_1(T) = -1$. We could also have gotten this result directly from formula (4.2.3):

$$\Delta_1(T) = \frac{v_2(4) - v_2(1)}{4-1} = \frac{1-4}{4-1} = -1.$$

For the sake of completeness, we consider finally the case where the first toss results in head. At time 1, we will have a portfolio valued at $X_1(H) = 0.40$. We choose

$$\Delta_1(H) = \frac{v_2(16) - v_2(4)}{16-4} = \frac{0-1}{16-4} = -\frac{1}{12}.$$

If the second toss results in head, at time 2 the value of our hedging portfolio is

$$X_2(HH) = 16\Delta_1(H) + \frac{5}{4}\big(0.40 - 8\Delta_1(H)\big) = 0 = v_2(16).$$

If the second toss results in tail, at time 2 the value of our hedging portfolio is

$$X_2(HT) = 4\Delta_1(H) + \frac{5}{4}\big(0.40 - 8\Delta_1(H)\big) = 1 = v_2(4).$$

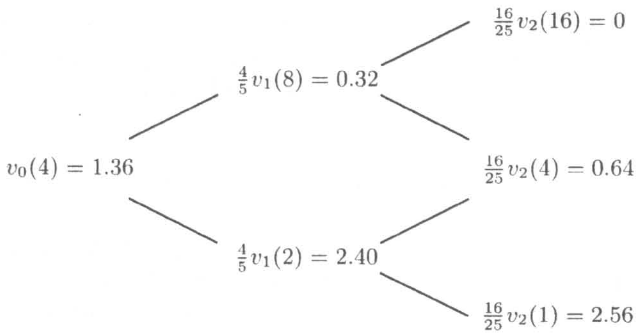

Fig. 4.2.3. Discounted American put prices.

Finally, we consider the discounted American put prices in Figure 4.2.3. These constitute a supermartingale under the risk-neutral probabilities $\tilde{p} = \tilde{q} = \frac{1}{2}$. At each node, the discounted American put price is greater than or equal to the average of the discounted prices at the two subsequent nodes. This price process is not a martingale because the inequality is strict at the time-one node corresponding to a tail on the first toss. □

The following theorem formalizes what we have seen in Example 4.2.1 and justifies the American algorithm (4.2.5), (4.2.6). We shall eventually prove the more general Theorems 4.4.3 and 4.4.4, which cover the case of path dependence as well as path independence, and thus do not pause to prove Theorem 4.2.2 below.

Theorem 4.2.2. (Replication of path-independent American derivatives) *Consider an N-period binomial asset-pricing model with $0 < d < 1 + r < u$ and with*

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

Let a payoff function $g(s)$ be given, and define recursively backward in time the sequence of functions $v_N(s), v_{N-1}(s), \dots, v_0(s)$ by (4.2.5), (4.2.6). Next define

$$\Delta_n = \frac{v_{n+1}(uS_n) - v_{n+1}(dS_n)}{(u-d)S_n}, \tag{4.2.7}$$

$$C_n = v_n(S_n) - \frac{1}{1+r}\left[\tilde{p}v_{n+1}(uS_n) + \tilde{q}v_{n+1}(dS_n)\right], \tag{4.2.8}$$

where n ranges between 0 and $N-1$. We have $C_n \geq 0$ for all n. If we set $X_0 = v_0(S_0)$ and define recursively forward in time the portfolio values $X_1, X_2, \dots, X_N$ by

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - C_n - \Delta_n S_n), \tag{4.2.9}$$

then we will have

$$X_n(\omega_1 \dots \omega_n) = v_n\big(S_n(\omega_1 \dots \omega_n)\big) \tag{4.2.10}$$

for all n and all $\omega_1 \dots \omega_n$. In particular, $X_n \geq g(S_n)$ for all n.

Equation (4.2.9) is the same as the wealth equation (1.2.14) of Chapter 1, except that we have included the possibility of consumption. Theorem 4.2.2 guarantees that we can hedge a short position in the American derivative security with intrinsic value $g(S_n)$ at each time n. In fact, we can do so and perhaps still consume at certain times. The value of our hedging portfolio X_n is always at least as great as the intrinsic value of the derivative security because of (4.2.10) and the fact, guaranteed by (4.2.6), that $v_n(S_n) \geq g(S_n)$. The nonnegativity of C_n also follows from (4.2.6), which implies that

$$v_n(S_n) \geq \frac{1}{1+r}\left[\tilde{p}v_{n+1}(uS_n) + \tilde{q}v_{n+1}(dS_n)\right].$$

4.3 Stopping Times

In general, the time at which an American derivative security should be exercised is random; it depends on the price movements of the underlying asset. We claimed in Example 4.2.1 that if the first coin toss results in tail, then the owner of the American put in that example should exercise at time one. On the other hand, if the first toss results in head, then the owner of the put should not exercise at time one but rather wait for the outcome of the second toss. Indeed, if the first toss results in head, then the stock price is $S_1(H) = 8$ and the put is out of the money. If the second toss results in another head, then $S_2(HH) = 16$, the put is still out of the money, and the owner should let it expire without exercising it. On the other hand, if the first toss is a head and the second toss is a tail, then $S_2(HT) = 4$, the put is in the money at time two, and the owner should exercise. We describe this exercise rule by the following random variable τ:

$$\tau(HH) = \infty, \quad \tau(HT) = 2, \quad \tau(TH) = 1, \quad \tau(TT) = 1, \tag{4.3.1}$$

which is displayed in Figure 4.3.1.

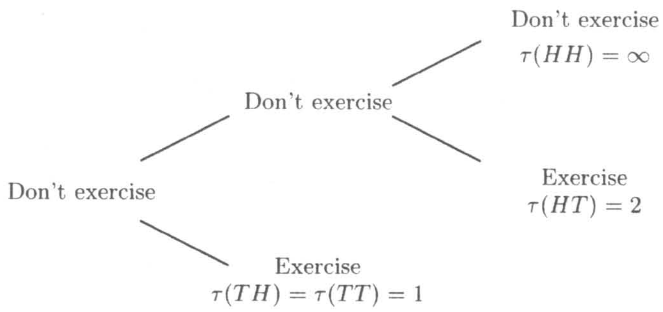

Fig. 4.3.1. Exercise rule τ.

On those paths where τ takes the value ∞, we mean that the option should be allowed to expire without exercise. In Example 4.2.1, there is only one such path, corresponding to HH. In the event the coin tosses result in HT, exercise should be done at time two. In the event the coin tosses result in TH or TT, then exercise should be done at time one. The random variable τ defined on $\Omega = \{HH, HT, TH, TT\}$ by (4.3.1) takes values in the set $\{0, 1, 2, \infty\}$. We can think of it as "stopping" the American put hedging problem by exercising the put, at least on three of the four sample points in Ω. It is a special case of a *stopping time* as defined in Definition 4.3.1 below.

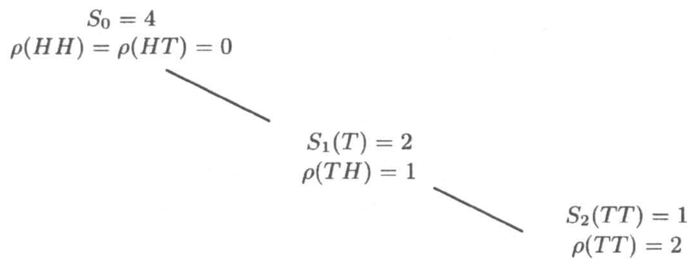

Fig. 4.3.2. Exercise rule ρ.

The owner of the put in this example will regret not exercising the put at time zero if the first coin toss results in a head. In particular, if she had foreknowledge of the coin tossing, she would rather use the exercise rule

$$\rho(HH) = 0, \quad \rho(HT) = 0, \quad \rho(TH) = 1, \quad \rho(TT) = 2, \tag{4.3.2}$$

which is displayed in Figure 4.3.2. If she could use this exercise rule, then regardless of the coin tossing, she would exercise the put in the money. The problem with the exercise rule ρ is that it cannot be implemented without "insider information." It calls for the decision of whether or not to exercise at time zero to be based on the outcome of the first coin toss. If the first coin toss results in a tail, then the decision of whether or not to exercise at time one is based on the outcome of the second coin toss. It is not a stopping time in the sense of Definition 4.3.1 below.

Definition 4.3.1. *In an N-period binomial model, a* stopping time *is a random variable τ that takes values $0, 1, \dots, N$ or ∞ and satisfies the condition that if $\tau(\omega_1\omega_2\dots\omega_n\omega_{n+1}\dots\omega_N) = n$, then $\tau(\omega_1\omega_2\dots\omega_n\omega'_{n+1}\dots\omega'_N) = n$ for all $\omega'_{n+1}\dots\omega'_N$.*

The condition in the definition above that if $\tau(\omega_1\omega_2\dots\omega_n\omega_{n+1}\dots\omega_N) = n$ then $\tau(\omega_1\omega_2\dots\omega_n\omega'_{n+1}\dots\omega'_N) = n$ for all $\omega'_{n+1}\dots\omega'_N$ ensures that stopping is based only on available information. If stopping occurs at time n, then this decision is based only on the first n coin tosses and not on the outcome of any subsequent toss.

Whenever we have a stochastic process and a stopping time, we can define a *stopped process* (see Figure 4.3.3). For example, let Y_n be the process of discounted American put prices in Figure 4.2.3; i.e.,

$$Y_0 = 1.36, \quad Y_1(H) = 0.32, \quad Y_1(T) = 2.40,$$
$$Y_2(HH) = 0, \quad Y_2(HT) = Y_2(TH) = 0.64, \quad Y_2(TT) = 2.56.$$

Let τ be the stopping time of (4.3.1). We define the stopped process $Y_{n\wedge\tau}$ by the formulas below. (The notation $n \wedge \tau$ denotes the minimum of n and τ.) We set

$$Y_{0\wedge\tau} = Y_0 = 1.36$$

because $0 \wedge \tau = 0$ regardless of the coin tossing. Similarly,

$$Y_{1\wedge\tau} = Y_1$$

because $1 \wedge \tau = 1$ regardless of the coin tossing. However, $2 \wedge \tau$ depends on the coin tossing; if we get HH or HT, then $2 \wedge \tau = 2$, but if we get TH or TT, we have $2 \wedge \tau = 1$. Therefore, we have the four cases

$$Y_{2\wedge\tau}(HH) = Y_2(HH) = 0,\ Y_{2\wedge\tau}(HT) = Y_2(HT) = 0.64,$$
$$Y_{2\wedge\tau}(TH) = Y_1(T) = 2.40,\ Y_{2\wedge\tau}(TT) = Y_1(T) = 2.40.$$

Note in this construction that the process continues on past time 1, even if τ takes the value 1. Time is not stopped. However, the *value* of the process is frozen at time τ. A better terminology might be to call $Y_{n\wedge\tau}$ a *frozen process*, but the term *stopped process* is already in universal use.

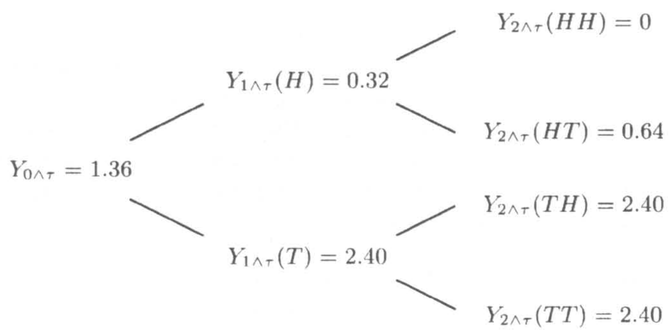

Fig. 4.3.3. A stopped process.

The discounted American put price process $Y_n = \left(\frac{4}{5}\right)^n v_n(S_n)$ in Figure 4.2.3 is a supermartingale but not a martingale under the risk-neutral probabilities $\tilde{p} = \tilde{q} = \frac{1}{2}$ because

$$2.40 = Y_1(T) > \frac{1}{2}Y_2(TH) + \frac{1}{2}Y_2(TT) = \frac{1}{2}\cdot 0.64 + \frac{1}{2}\cdot 2.56 = 1.60.$$

The stopped process $Y_{n\wedge\tau}$ is a martingale. In particular,

$$2.40 = Y_{1\wedge\tau}(T) = \frac{1}{2}Y_{2\wedge\tau}(TH) + \frac{1}{2}Y_{2\wedge\tau}(TT) = \frac{1}{2}\cdot 2.40 + \frac{1}{2}\cdot 2.40.$$

This observation is true generally. Under the risk-neutral probabilities, a discounted American derivative security price process is a supermartingale. However, if this process is stopped at the optimal exercise time, it becomes a martingale. If the owner of the security permits a time to pass in which the supermartingale inequality is strict, she has failed to exercise optimally.

We have just seen by example that we can stop a process that is not a martingale and thereby obtain a martingale. Of course, if we stop a process that is already a martingale, we will obtain a process that may be different but is still a martingale. This trivial observation is one consequence of a general theorem about stopping times called the Optional Sampling Theorem.

Theorem 4.3.2 (Optional sampling—Part I). *A martingale stopped at a stopping time is a martingale. A supermartingale (or submartingale) stopped at a stopping time is a supermartingale (or submartingale, respectively).*

We illustrate the first statement of the theorem with an example. In Figure 4.3.4, we consider the discounted stock price process $M_n = \left(\frac{4}{5}\right)^n S_n$, which is a martingale under the risk-neutral probabilities $\tilde{p} = \tilde{q} = \frac{1}{2}$. At every node, the value is the average of the values at the two subsequent nodes. In Figure 4.3.5, we show the same process stopped by the stopping time τ of (4.3.1). Again, at every node, the value is the average of the values at the two subsequent nodes.

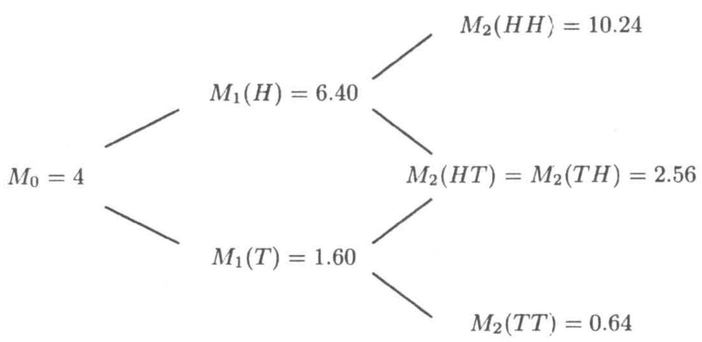

Fig. 4.3.4. Discounted stock price.

Figure 4.3.3 shows a supermartingale stopped at a stopping time, and the resulting process is a martingale, which is of course still a supermartingale. This illustrates the second statement in Theorem 4.3.2.

Finally, we note that if we stop the discounted stock price process of Figure 4.3.4 using the random time ρ of (4.3.2), which is not a stopping time, we destroy the martingale property. Figure 4.3.6 shows the stopped process. The

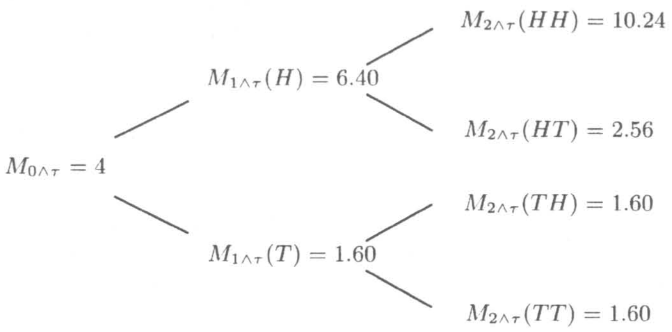

Fig. 4.3.5. Discounted stock price stopped at a stopping time.

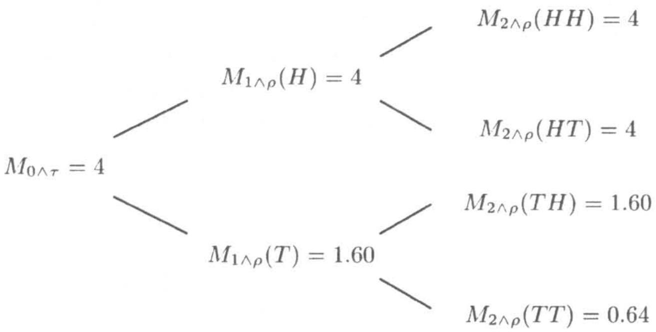

Fig. 4.3.6. Discounted stock price stopped at a nonstopping time.

random time ρ looks ahead and decides to stop if the stock price is about to go up. This introduces a downward bias in the discounted stock price.

A submartingale has a tendency to go up. In particular, if X_n is a submartingale, then $\mathbb{E}X_m \leq \mathbb{E}X_n$ whenever $m \leq n$. This inequality still holds if we replace m by $\tau \wedge n$, where τ is a stopping time.

Theorem 4.3.3 (Optional sampling—Part II). *Let X_n, $n = 0, 1, \ldots, N$ be a submartingale, and let τ be a stopping time. Then $\mathbb{E}X_{n\wedge\tau} \leq \mathbb{E}X_n$. If X_n is a supermartingale, then $\mathbb{E}X_{n\wedge\tau} \geq \mathbb{E}X_n$; if X_n is a martingale, then $\mathbb{E}X_{n\wedge\tau} = \mathbb{E}X_n$.*

The expectation in Theorem 4.3.3 is computed under the probabilities that make X_n be a submartingale (or supermartingale or martingale). In particular, if X_n is a submartingale under the risk-neutral probabilities in a binomial model, then the conclusion of the theorem would be $\widetilde{\mathbb{E}}X_{n\wedge\tau} \leq \widetilde{\mathbb{E}}X_n$.

4.4 General American Derivatives

In this section, we introduce American derivative securities whose intrinsic value is permitted to be path-dependent. We define the price process for such a security and develop its properties. We also show how to hedge a short position in such a derivative security and study the optimal exercise time. All claims made in this section are supported by mathematical proofs.

We work within the context of an N-period binomial model with up factor u, down factor d, and interest rate r satisfying the no-arbitrage condition $0 < d < 1 + r < u$. In such a model, we define $\mathcal{S}_n$ to be the set of all stopping times τ that take values in the set $\{n, n+1, \ldots, N, \infty\}$. In particular, the set $\mathcal{S}_0$ contains every stopping time. A stopping time in $\mathcal{S}_N$ can take the value N on some paths, the value ∞ on others, and can take no other value.

Definition 4.4.1. *For each n, $n = 0, 1, \ldots, N$, let G_n be a random variable depending on the first n coin tosses. An* American derivative security *with* intrinsic value process *G_n is a contract that can be exercised at any time prior to and including time N and, if exercised at time n, pays off G_n. We define the* price process *V_n for this contract by the* American risk-neutral pricing formula

$$V_n = \max_{\tau \in \mathcal{S}_n} \widetilde{\mathbb{E}}_n \left[\mathbb{I}_{\{\tau \leq N\}} \frac{1}{(1+r)^{\tau-n}} G_\tau \right], \quad n = 0, 1, \ldots, N. \tag{4.4.1}$$

The idea behind (4.4.1) is the following. Suppose the American derivative security is not exercised at times $0, 1, \ldots, n-1$ and we are trying to determine its value at time n. At time n, the owner of the derivative can choose to exercise it immediately or postpone exercise to some later date. The date at which she exercises, if she does exercise, can depend on the path of the stock price up to the exercise time but not beyond it. In other words, the exercise date will be a stopping time τ. Since exercise was not done before time n, this stopping time must be in $\mathcal{S}_n$. Of course, if she never exercises ($\tau = \infty$), then she receives zero payoff. The term $\mathbb{I}_{\{\tau \leq N\}}$ appears in (4.4.1) to tell us that $\mathbb{I}_{\{\tau\leq N\}} \frac{1}{(1+r)^{\tau-n}} G_\tau$ should be replaced by zero on those paths for which $\tau = \infty$. When the owner exercises according to a stopping time $\tau \in \mathcal{S}_n$, the value of the derivative to her at time n is the risk-neutral discounted expectation of its payoff. She should choose τ to make this as large as possible.

One of the immediate consequences of this definition is that

$$V_N = \max\{G_N, 0\}. \tag{4.4.2}$$

To see that, we take $n = N$ in (4.4.1) so that it becomes

$$V_N = \sup_{\tau \in \mathcal{S}_N} \mathbb{I}_{\{\tau \leq N\}} \frac{1}{(1+r)^{\tau-N}} G_\tau.$$

A stopping time in $\mathcal{S}_N$ takes only the values N and ∞, and for such a stopping time

$$\mathbb{I}_{\{\tau \le N\}} \frac{1}{(1+r)^{\tau - N}} G_\tau = \mathbb{I}_{\{\tau = N\}} G_N.$$

In order to make this as large as possible, we should choose $\tau(\omega_1 \ldots \omega_N) = N$ if $G_N(\omega_1 \ldots \omega_N) > 0$ and $\tau(\omega_1 \ldots \omega_N) = \infty$ if $G_N(\omega_1 \ldots \omega_N) \le 0$. With this choice of τ, we have $\mathbb{I}_{\{\tau=N\}} G_N = \max\{G_N, 0\}$, and (4.4.2) is established.

Before working out other consequences of Definition 4.4.1, we rework Example 4.2.1 to verify that this definition is consistent with the American put prices obtained in that example.

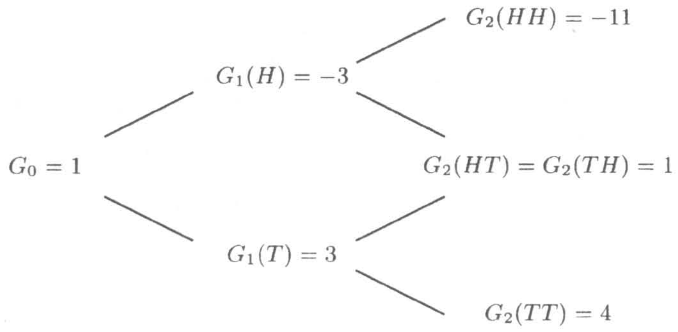

Fig. 4.4.1. Intrinsic value.

Example 4.2.1 continued For the American put with strike 5 on the stock given in Figure 4.2.1, the intrinsic value is given in Figure 4.4.1. In this example, $N = 2$ and (4.4.2) becomes

$$V_2(HH) = 0,\ V_2(HT) = V_2(TH) = 1,\ V_2(TT) = 4.$$

We next apply (4.4.1) with $n = 1$, first considering the case of head on the first toss. Then

$$V_1(H) = \max_{\tau \in \mathcal{S}_1} \widetilde{\mathbb{E}}_1 \left[\mathbb{I}_{\{\tau \le 2\}} \left(\frac{4}{5}\right)^{\tau - 1} G_\tau \right](H). \tag{4.4.3}$$

To make the conditional expectation on the right-hand side of (4.4.3) as large as possible, we should take $\tau(HH) = \infty$ (do not exercise in the case of HH) and take $\tau(HT) = 2$ (exercise at time one in the case of HT). The decision of whether or not to exercise at time two is based on the information available at time two, so this does not violate the property required of stopping times. This exercise policy makes (4.4.3) equal to

$$V_1(H) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \left(\frac{4}{5}\right)^{2-1} G_2(HT) = 0.40.$$

We next apply (4.4.1) with $n = 1$ when the first toss results in a tail. In this case,

$$V_1(T) = \max_{\tau \in \mathcal{S}_1} \widetilde{\mathbb{E}}_1 \left[\mathbb{I}_{\{\tau \le 2\}} \left(\frac{4}{5}\right)^{\tau - 1} G_\tau \right](T). \tag{4.4.4}$$

To make this conditional expectation as large as possible, knowing that the first toss results in a tail, we must consider two possibilities: exercise at time one or exercise at time two. It is clear that we want to exercise at one of these two times because the option is in the money regardless of the second coin toss. If we take $\tau(TH) = \tau(TT) = 1$, then

$$\widetilde{\mathbb{E}}_1 \left[\mathbb{I}_{\{\tau \le 2\}} \left(\frac{4}{5}\right)^{\tau - 1} G_\tau \right](T) = G_1(T) = 3.$$

If we take $\tau(TH) = \tau(TT) = 2$, then

$$\widetilde{\mathbb{E}}_1 \left[\mathbb{I}_{\{\tau \le 2\}} \left(\frac{4}{5}\right)^{\tau - 1} G_\tau \right](T) = \frac{4}{5}\left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 4\right) = 2.$$

We cannot choose $\tau(TH) = 1$ and $\tau(TT) = 2$ because that would violate the property of stopping times. We see then that (4.4.4) yields $V_1(T) = 3$.

Finally, when $n = 0$, we have

$$V_0 = \max_{\tau \in \mathcal{S}_0} \widetilde{\mathbb{E}} \left[\mathbb{I}_{\{\tau \le 2\}} \left(\frac{4}{5}\right)^{\tau} G_\tau \right]. \tag{4.4.5}$$

There are many stopping times to consider (see Exercise 4.5), but a moment's reflection shows that the one that makes $\widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau \le 2\}} \left(\frac{4}{5}\right)^{\tau} G_\tau\right]$ as large as possible is

$$\tau(HH) = \infty,\ \tau(HT) = 2,\ \tau(TH) = \tau(TT) = 1. \tag{4.4.6}$$

With this stopping time, (4.4.5) becomes

$$V_0 = \frac{1}{4} \cdot 0 + \frac{1}{4}\left(\frac{4}{5}\right)^2 G_2(HT) + \frac{1}{2} \cdot \frac{4}{5} \cdot G_1(T) = \frac{1}{4} \cdot \frac{16}{25} \cdot 1 + \frac{1}{2} \cdot \frac{4}{5} \cdot 3 = 1.36. \tag{4.4.7}$$

We record the option prices in Figure 4.4.2. These agree with the prices in Figure 4.2.2, the only difference being that in Figure 4.2.2 these prices are recorded as functions v_n of the underlying stock prices and here they are recorded as random variables (i.e., functions of the coin tosses). The prices in the two figures are related by the formula $V_n = v_n(S_n)$. □

We now develop the properties of the American derivative security price process of Definition 4.4.1. These properties justify calling V_n defined by (4.4.1) the *price* of the derivative security.

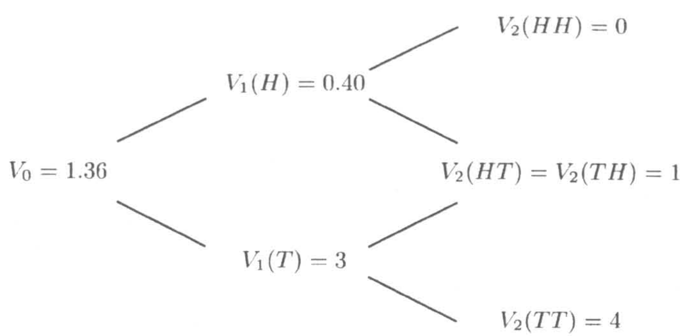

Fig. 4.4.2. American put prices.

Theorem 4.4.2. *The American derivative security price process given by Definition 4.4.1 has the following properties:*

(i) $V_n \geq \max\{G_n, 0\}$ *for all* n*;*
(ii) the discounted process $\frac{1}{(1+r)^n}V_n$ *is a supermartingale;*
(iii) if Y_n *is another process satisfying* $Y_n \geq \max\{G_n, 0\}$ *for all* n *and for which* $\frac{1}{(1+r)^n}Y_n$ *is a supermartingale, then* $Y_n \geq V_n$ *for all* n*.*

We summarize property (iii) by saying that V_n *is the* smallest *process satisfying (i) and (ii).*

We shall see that property (ii) in Theorem 4.4.2 guarantees that an agent beginning with initial capital V_0 can construct a hedging portfolio whose value at each time n is V_n. Property (i) guarantees that if an agent does this, he has hedged a short position in the derivative security; no matter when it is exercised, the agent's hedging portfolio value is sufficient to pay off the derivative security. Thus, (i) and (ii) guarantee that the derivative security price is acceptable to the seller. Condition (iii) says that the price is no higher than necessary in order to be acceptable to the seller. This condition ensures that the price is fair for the buyer.

PROOF: We first establish (i). Let n be given, and consider the stopping time $\widehat{\tau}$ in $\mathcal{S}_n$ that takes the value n, regardless of the coin tossing. Then

$$\widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\widehat{\tau}\leq N\}}\frac{1}{(1+r)^{\widehat{\tau}-n}}G_\tau\right] = G_n.$$

Since V_n is the largest possible value we can obtain for $\widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau\leq N\}}\frac{1}{(1+r)^{\tau-n}}G_\tau\right]$ when we consider all stopping times $\tau \in \mathcal{S}_n$, we must have $V_n \geq G_n$. On the other hand, if we take $\overline{\tau}$ to be the stopping time in $\mathcal{S}_n$ that takes the value ∞, regardless of the coin tossing, then

$$\widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\overline{\tau}\leq N\}}\frac{1}{(1+r)^{\overline{\tau}-n}}G_{\overline{\tau}}\right] = 0.$$

Again, V_n is the maximum of expressions of this type, and hence $V_n \geq 0$. We conclude that (i) holds.

We next prove (ii). Let n be given, and suppose τ^* attains the maximum in the definition of V_{n+1}; i.e., $\tau^* \in \mathcal{S}_{n+1}$ and

$$V_{n+1} = \widetilde{\mathbb{E}}_{n+1}\left[\mathbb{I}_{\{\tau^*\leq N\}}\frac{1}{(1+r)^{\tau^*-n-1}}G_{\tau^*}\right]. \tag{4.4.8}$$

But $\tau^* \in \mathcal{S}_n$ also, which together with iterated conditioning implies

$$\begin{aligned}
V_n &\geq \widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau^*\leq N\}}\frac{1}{(1+r)^{\tau^*-n}}G_{\tau^*}\right]\\
&= \widetilde{\mathbb{E}}_n\left[\widetilde{\mathbb{E}}_{n+1}\left[\mathbb{I}_{\{\tau^*\leq N\}}\frac{1}{(1+r)^{\tau^*-n}}G_{\tau^*}\right]\right]\\
&= \widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}\widetilde{\mathbb{E}}_{n+1}\left[\mathbb{I}_{\{\tau^*\leq N\}}\frac{1}{(1+r)^{\tau^*-n-1}}G_{\tau^*}\right]\right]\\
&= \widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}V_{n+1}\right].
\end{aligned} \tag{4.4.9}$$

Dividing both sides by $(1+r)^n$, we obtain the supermartingale property for the discounted price process:

$$\frac{1}{(1+r)^n}V_n \geq \widetilde{\mathbb{E}}_n\left[\frac{1}{(1+r)^{n+1}}V_{n+1}\right].$$

Finally, we prove (iii). Let Y_n be another process satisfying conditions (i) and (ii). Let $n \leq N$ be given and let τ be a stopping time in $\mathcal{S}_n$. Because $Y_k \geq \max\{G_k, 0\}$ for all k, we have

$$\begin{aligned}
\mathbb{I}_{\{\tau\leq N\}}G_\tau &\leq \mathbb{I}_{\{\tau\leq N\}}\max\{G_\tau, 0\}\\
&\leq \mathbb{I}_{\{\tau\leq N\}}\max\{G_{N\wedge\tau}, 0\} + \mathbb{I}_{\{\tau=\infty\}}\max\{G_{N\wedge\tau}, 0\}\\
&= \max\{G_{N\wedge\tau}, 0\}\\
&\leq Y_{N\wedge\tau}.
\end{aligned}$$

We next use the Optional Sampling Theorem 4.3.2 and the supermartingale property for $\frac{1}{(1+r)^k}Y_k$ to write

$$\begin{aligned}
\widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau\leq N\}}\frac{1}{(1+r)^\tau}G_\tau\right] &= \widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau\leq N\}}\frac{1}{(1+r)^{N\wedge\tau}}G_\tau\right]\\
&\leq \widetilde{\mathbb{E}}_n\left[\frac{1}{(1+r)^{N\wedge\tau}}Y_{N\wedge\tau}\right]\\
&\leq \frac{1}{(1+r)^{n\wedge\tau}}Y_{n\wedge\tau}\\
&= \frac{1}{(1+r)^n}Y_n,
\end{aligned}$$

the equality at the end being a consequence of the fact that $\tau \in \mathcal{S}_n$ is greater than or equal to n on every path. Multiplying by $(1+r)^n$, we obtain

$$\widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau\le N\}}\frac{1}{(1+r)^{\tau-n}}G_\tau\right]\le Y_n.$$

Since V_n is the maximum value we can obtain for $\widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau\le N\}}\frac{1}{(1+r)^{\tau-n}}G_\tau\right]$ as τ ranges over $\mathcal{S}_n$, and all these values are less than or equal to Y_n, we must have $V_n \le Y_n$. □

We now generalize the American pricing algorithm given by (4.2.5) and (4.2.6) to path-dependent securities

Theorem 4.4.3. *We have the following American pricing algorithm for the path-dependent derivative security price process given by Definition 4.4.1:*

$$V_N(\omega_1\dots\omega_N)=\max\{G_N(\omega_1\dots\omega_N),0\}, \tag{4.4.10}$$

$$V_n(\omega_1\dots\omega_n)=\max\left\{G_n(\omega_1\dots\omega_n),\ \frac{1}{1+r}\left[\tilde{p}V_{n+1}(\omega_1\dots\omega_nH)+\tilde{q}V_{n+1}(\omega_1\dots\omega_nT)\right]\right\}, \tag{4.4.11}$$

for $n=N-1,\dots,0$.

PROOF: We shall prove that V_n defined recursively by (4.4.10) and (4.4.11) satisfies properties (i) and (ii) of Theorem 4.4.2 and is the smallest process with these properties. According to Theorem 4.4.2, the process V_n given by (4.4.1) is the smallest process with these properties, and hence the algorithm (4.4.10), (4.4.11) must generate the same process as formula (4.4.1).

We first establish property (i) of Theorem 4.4.2. It is clear that V_N defined by (4.4.10) satisfies property (i) with $n=N$. We proceed by induction backward in time. Suppose that for some n between 0 and $N-1$ we have $V_{n+1}\ge\max\{G_{n+1},0\}$. Then, from (4.4.11), we see that

$$V_n(\omega_1\dots\omega_n)\ge\max\left\{G_n(\omega_1\dots\omega_n),0\right\}.$$

This completes the induction step and shows that V_n defined recursively by (4.4.10), (4.4.11) satisfies property (i) of Theorem 4.4.2.

We next verify that $\frac{1}{(1+r)^n}V_n$ is a supermartingale. From (4.4.11), we see immediately that

$$\begin{aligned}V_n(\omega_1\dots\omega_n)&\ge\frac{1}{1+r}\left[\tilde{p}V_{n+1}(\omega_1\dots\omega_nH)+\tilde{q}V_{n+1}(\omega_1\dots\omega_nT)\right]\\&=\widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}V_{n+1}\right](\omega_1\dots\omega_n).\end{aligned}\tag{4.4.12}$$

Multiplying both sides by $\frac{1}{(1+r)^n}$, we obtain the desired supermartingale property.

Finally, we must show that V_n defined by (4.4.10), (4.4.11) and satisfying (i) and (ii) of Theorem 4.4.2 is the smallest process satisfying (i) and (ii). It is immediately clear from (4.4.10) that V_N is the smallest random variable satisfying $V_N\ge\max\{G_N,0\}$. We proceed by induction backward in time. Suppose that, for some n between 0 and $N-1$, V_{n+1} is as small as possible. The supermartingale property (ii) implies that V_n must satisfy (4.4.12). In order to satisfy property (i), V_n must be greater than or equal to G_n. Therefore, properties (i) and (ii) of Theorem 4.4.2 imply

$$V_n(\omega_1\dots\omega_n)\ge\max\left\{G_n(\omega_1\dots\omega_n),\ \frac{1}{1+r}\left[\tilde{p}V_{n+1}(\omega_1\dots\omega_nH)+\tilde{q}V_{n+1}(\omega_1\dots\omega_nT)\right]\right\} \tag{4.4.13}$$

for $n=N-1,\dots,0$. But (4.4.11) defines $V_n(\omega_1\dots\omega_n)$ to be equal to the right-hand side of (4.4.13), which means $V_n(\omega_1\dots\omega_n)$ is as small as possible. □

In order to justify Definition 4.4.1 for American derivative security prices, we must show that a short position can be hedged using these prices. This requires a generalization of Theorem 4.2.2 to the path-dependent case.

Theorem 4.4.4 (Replication of path-dependent American derivatives). *Consider an N-period binomial asset-pricing model with $0<d<1+r<u$ and with*

$$\tilde{p}=\frac{1+r-d}{u-d},\quad \tilde{q}=\frac{u-1-r}{u-d}.$$

For each n, $n=0,1,\dots,N$, let G_n be a random variable depending on the first n coin tosses. With V_n, $n=0,1,\dots,N$, given by Definition 4.4.1, we define

$$\Delta_n(\omega_1\dots\omega_n)=\frac{V_{n+1}(\omega_1\dots\omega_nH)-V_{n+1}(\omega_1\dots\omega_nT)}{S_{n+1}(\omega_1\dots\omega_nH)-S_{n+1}(\omega_1\dots\omega_nT)}, \tag{4.4.14}$$

$$C_n(\omega_1\dots\omega_n)=V_n(\omega_1\dots\omega_n)-\frac{1}{1+r}\left[\tilde{p}V_{n+1}(\omega_1\dots\omega_nH)+\tilde{q}V_{n+1}(\omega_1\dots\omega_nT)\right], \tag{4.4.15}$$

where n ranges between 0 and $N-1$. We have $C_n\ge0$ for all n. If we set $X_0=V_0$ and define recursively forward in time the portfolio values $X_1,X_2,\dots,X_N$ by

$$X_{n+1}=\Delta_nS_{n+1}+(1+r)(X_n-C_n-\Delta_nS_n), \tag{4.4.16}$$

then we have

$$X_n(\omega_1\dots\omega_n)=V_n(\omega_1\dots\omega_n) \tag{4.4.17}$$

for all n and all $\omega_1\dots\omega_n$. In particular, $X_n\ge G_n$ for all n.

PROOF: The nonnegativity of C_n is a consequence of property (ii) of Theorem 4.4.2 or, equivalently, (4.4.12).

To prove (4.4.17), we proceed by induction on n. This part of the proof is the same as the proof of Theorem 2.4.8. The induction hypothesis is that $X_n(\omega_1 \dots \omega_n) = V_n(\omega_1 \dots \omega_n)$ for some $n \in \{0, 1, \dots, N-1\}$ and all $\omega_1 \dots \omega_n$. We need to show that

$$X_{n+1}(\omega_1 \dots \omega_n H) = V_{n+1}(\omega_1 \dots \omega_n H), \tag{4.4.18}$$

$$X_{n+1}(\omega_1 \dots \omega_n T) = V_{n+1}(\omega_1 \dots \omega_n T). \tag{4.4.19}$$

We prove (4.4.18); the proof of (4.4.19) is analogous.

Note first that

$$\begin{aligned} &V_n(\omega_1 \dots \omega_n) - C_n(\omega_1 \dots \omega_n) \\ &= \frac{1}{1+r}\Big[\tilde{p}V_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}V_{n+1}(\omega_1 \dots \omega_n T)\Big]. \end{aligned}$$

Since $\omega_1 \dots \omega_n$ will be fixed for the rest of the proof, we will suppress these symbols. For example, the last equation will be written simply as

$$V_n - C_n = \frac{1}{1+r}\Big[\tilde{p}V_{n+1}(H) + \tilde{q}V_{n+1}(T)\Big].$$

We compute

$$\begin{aligned} X_{n+1}(H) &= \Delta_n S_{n+1}(H) + (1+r)(X_n - C_n - \Delta_n S_n) \\ &= \frac{V_{n+1}(H) - V_{n+1}(T)}{S_{n+1}(H) - S_{n+1}(T)}\big(S_{n+1}(H) - (1+r)S_n\big) \\ &\qquad +(1+r)(V_n - C_n) \\ &= \frac{V_{n+1}(H) - V_{n+1}(T)}{(u-d)S_n}\big(uS_n - (1+r)S_n\big) \\ &\qquad +\tilde{p}V_{n+1}(H) + \tilde{q}V_{n+1}(T) \\ &= \big(V_{n+1}(H) - V_{n+1}(T)\big)\frac{u-1-r}{u-d} + \tilde{p}V_{n+1}(H) + \tilde{q}V_{n+1}(T) \\ &= \big(V_{n+1}(H) - V_{n+1}(T)\big)\tilde{q} + \tilde{p}V_{n+1}(H) + \tilde{q}V_{n+1}(T) \\ &= (\tilde{p} + \tilde{q})V_{n+1}(H) \;=\; V_{n+1}(H). \end{aligned}$$

This is (4.4.18).

The final claim of the theorem, that $X_n \geq G_n$ for all n, follows from (4.4.17) and property (i) of Theorem 4.4.2. □

Theorem 4.4.4 shows that the American derivative security price given by (4.4.1) is acceptable to the seller because he can construct a hedge for the short position. We next argue that it is also acceptable to the buyer. Let us fix n, imagine we have gotten to time n without the derivative security being exercised, and denote by $\tau^* \in \mathcal{S}_n$ the stopping time that attains the maximum in (4.4.1), so that

$$V_n = \widetilde{\mathbb{E}}_n\left[\mathbb{I}_{\{\tau^* \leq N\}}\frac{1}{(1+r)^{\tau^*-n}}G_{\tau^*}\right]. \tag{4.4.20}$$

For $k = n, n+1, \dots, N$, define

$$C_k = \mathbb{I}_{\{\tau^*=k\}}G_k.$$

If the owner of the derivative security exercises it according to the stopping time τ^*, then she will receive the cash flows $C_n, C_{n+1}, \dots, C_N$ at times $n, n+1, \dots, N$, respectively. Actually, at most one of these C_k values is non-zero. If the option is exercised at or before the expiration time N, then the C_k corresponding to the exercise time is the only nonzero payment among them. However, on different paths, this payment comes at different times. In any case, (4.4.20) becomes

$$V_n = \widetilde{\mathbb{E}}_n\left[\sum_{k=n}^{N}\mathbb{I}_{\{\tau^*=k\}}\frac{1}{(1+r)^{k-n}}G_k\right] = \widetilde{\mathbb{E}}_n\left[\sum_{k=n}^{N}\frac{C_k}{(1+r)^{k-n}}\right].$$

We saw in Theorem 2.4.8 of Chapter 2 that this is just the value at time n of the cash flows $C_n, C_{n+1}, \dots, C_N$, received at times $n, n+1, \dots, N$, respectively. Once the option holder decides on the exercise strategy τ^*, this is exactly the contract she holds. Thus, the American derivative security price V_n is acceptable to her.

It remains to provide a method for the American derivative security owner to choose an optimal exercise time. We shall consider this problem with $n = 0$ (i.e., seek a stopping time $\tau^* \in \mathcal{S}_0$ that achieves the maximum in (4.4.1) when $n = 0$).

Theorem 4.4.5 (Optimal exercise). *The stopping time*

$$\tau^* = \min\{n; V_n = G_n\} \tag{4.4.21}$$

maximizes the right-hand side of (4.4.1) when $n = 0$; i.e.,

$$V_0 = \widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau^* \leq N\}}\frac{1}{(1+r)^{\tau^*}}G_{\tau^*}\right]. \tag{4.4.22}$$

The value of an American derivative security is always greater than or equal to its intrinsic value. The stopping time τ^* of (4.4.21) is the first time these two are equal. In may be that they are never equal. For example, the value of an American put is always greater than or equal to zero, but the put can always be out of the money (i.e., with negative intrinsic value). In this case, the minimum in (4.4.21) is over the empty set (the set of integers n for which $V_n = G_n$ is the empty set), and we follow the mathematical convention that the minimum over the empty set is ∞. For us, $\tau^* = \infty$ is synonymous with the derivative security expiring unexercised.

PROOF OF THEOREM 4.4.5: We first observe that the stopped process

$$\frac{1}{(1+r)^{n\wedge\tau^*}} V_{n\wedge\tau^*} \tag{4.4.23}$$

is a martingale under the risk-neutral probability measure. This is a consequence of (4.4.11). Indeed, if the first n coin tosses result in $\omega_1 \dots \omega_n$ and along this path $\tau^* \geq n+1$, then we know that $V_n(\omega_1 \dots \omega_n) > G_n(\omega_1 \dots \omega_n)$ and (4.4.11) implies

$$\begin{aligned} &V_{n\wedge\tau^*}(\omega_1 \dots \omega_n) \\ &= V_n(\omega_1 \dots \omega_n) \\ &= \frac{1}{1+r}\Big[\tilde{p}V_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}V_{n+1}(\omega_1 \dots \omega_n T)\Big] \\ &= \frac{1}{1+r}\Big[\tilde{p}V_{(n+1)\wedge\tau^*}(\omega_1 \dots \omega_n H) + \tilde{q}V_{(n+1)\wedge\tau^*}(\omega_1 \dots \omega_n T)\Big]. \end{aligned}$$

This is the martingale property for the process (4.4.23). On the other hand, if along the path $\omega_1 \dots \omega_n$ we have $\tau^* \leq n$, then

$$\begin{aligned} &V_{n\wedge\tau^*}(\omega_1 \dots \omega_n) \\ &= V_{\tau^*}(\omega_1 \dots \omega_{\tau^*}) \\ &= \tilde{p}V_{\tau^*}(\omega_1 \dots \omega_{\tau^*}) + \tilde{q}V_{\tau^*}(\omega_1 \dots \omega_{\tau^*}) \\ &= \tilde{p}V_{(n+1)\wedge\tau^*}(\omega_1 \dots \omega_n H) + \tilde{q}V_{(n+1)\wedge\tau^*}(\omega_1 \dots \omega_n T). \end{aligned}$$

Again we have the martingale property.

Since the stopped process (4.4.23) is a martingale, we have

$$\begin{aligned} V_0 &= \widetilde{\mathbb{E}}\left[\frac{1}{(1+r)^{N\wedge\tau^*}} V_{N\wedge\tau^*}\right] \\ &= \widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau^*\leq N\}} \frac{1}{(1+r)^{\tau^*}} G_{\tau^*}\right] + \widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau^*=\infty\}} \frac{1}{(1+r)^N} V_N\right]. \end{aligned} \tag{4.4.24}$$

But on those paths for which $\tau^* = \infty$, we must have $V_n > G_n$ for all n and, in particular, $V_N > G_N$. In light of (4.4.10), this can only happen if $G_N < 0$ and $V_N = 0$. Therefore, $\mathbb{I}_{\{\tau^*=\infty\}} V_N = 0$ and (4.4.24) can be simplified to

$$V_0 = \widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau^*\leq N\}} \frac{1}{(1+r)^{\tau^*}} G_{\tau^*}\right]. \tag{4.4.25}$$

This is (4.4.22). □

4.5 American Call Options

We saw in Example 4.2.1 that it is sometimes optimal to exercise an American put "early" (i.e., before expiration). For an American call on a non-dividend-paying stock, there is no advantage to early exercise. This is a consequence of Jensen's inequality for conditional expectations, as we show below.

Let $g : [0, \infty) \to \mathbb{R}$ be a convex function satisfying $g(0) = 0$ (see Figure 4.5.1). This means that whenever $s_1 \geq 0$, $s_2 \geq 0$, and $0 \leq \lambda \leq 1$, we have

$$g\big(\lambda s_1 + (1-\lambda)s_2\big) \leq \lambda g(s_1) + (1-\lambda)g(s_2). \tag{4.5.1}$$

For instance, we might have $g(s) = (s-K)^+$, the payoff of a call with strike K.

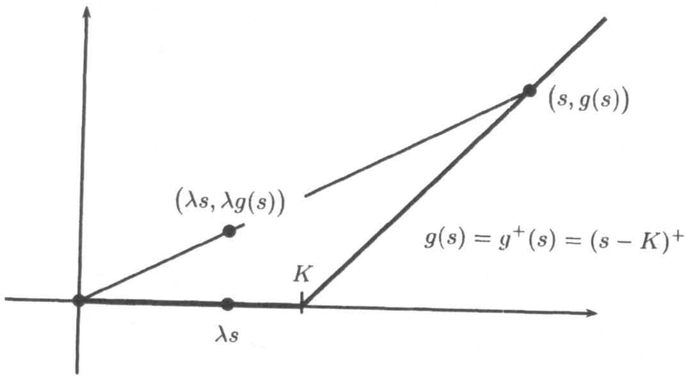

Fig. 4.5.1. Convex function with $g(0) = 0$.

Theorem 4.5.1. *Consider an N-period binomial asset-pricing model with $0 < d < 1 + r < u$ and $r \geq 0$. In this model, consider an American derivative security with convex payoff function $g(s)$ satisfying $g(0) = 0$. The value of this derivative security at time zero, which is (see Definition 4.4.1)*

$$V_0^A = \max_{\tau\in\mathcal{S}_0} \widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau\leq N\}} \frac{1}{(1+r)^\tau} g(S_\tau)\right], \tag{4.5.2}$$

is the same as the value of the European derivative security with payoff $g(S_N)$ at expiration N, which is (see Theorem 2.4.7 of Chapter 2)

$$V_0^E = \widetilde{\mathbb{E}}\left[\frac{1}{(1+r)^N} \max\{g(S_N), 0\}\right]. \tag{4.5.3}$$

PROOF: Under our assumptions, g can take negative values. We therefore introduce the function

$$g^+(s) = \max\{g(s), 0\},$$

which takes only nonnegative values and satisfies $g^+(0) = 0$. Furthermore, g^+ is convex. To see this, observe that because g satisfies (4.5.1), we have

$$g(\lambda s_1 + (1-\lambda)s_2) \le \lambda g^+(s_1) + (1-\lambda)g^+(s_2) \text{ for all } s_1 \ge 0,\ s_2 \ge 0,\ \lambda \in [0,1].$$

We also have $0 \le \lambda g^+(s_1) + (1-\lambda)g^+(s_2)$, and hence

$$g^+(\lambda s_1 + (1-\lambda)s_2) = \max\left\{0, g(\lambda s_1 + (1-\lambda)s_2)\right\} \le \lambda g^+(s_1) + (1-\lambda)g^+(s_2).$$

This last inequality establishes the convexity of g^+. Taking $s_1 = s$ and $s_2 = 0$ in this inequality, we see that

$$g^+(\lambda s) \le \lambda g^+(s) \text{ for all } s \ge 0,\ \lambda \in [0,1]. \tag{4.5.4}$$

Because $\frac{1}{(1+r)^n} S_n$ is a martingale under the risk-neutral probabilities, we have $S_n = \widetilde{\mathbb{E}}_n\left[\frac{1}{1+r} S_{n+1}\right]$ and

$$g^+(S_n) = g^+\left(\widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}S_{n+1}\right]\right).$$

The conditional Jensen's inequality of Theorem 2.3.2(v) of Chapter 2 implies that

$$g^+\left(\widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}S_{n+1}\right]\right) \le \widetilde{\mathbb{E}}_n\left[g^+\left(\frac{1}{1+r}S_{n+1}\right)\right]. \tag{4.5.5}$$

Taking $\lambda = \frac{1}{1+r}$ in (4.5.4), we obtain

$$\widetilde{\mathbb{E}}_n\left[g^+\left(\frac{1}{1+r}S_{n+1}\right)\right] \le \widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}g^+(S_{n+1})\right]. \tag{4.5.6}$$

Putting all of this together, we see that

$$g^+(S_n) \le \widetilde{\mathbb{E}}_n\left[\frac{1}{1+r}g^+(S_{n+1})\right],$$

and multiplication of both sides by $\frac{1}{(1+r)^n}$ yields the submartingale property

$$\frac{1}{(1+r)^n}g^+(S_n) \le \widetilde{\mathbb{E}}_n\left[\frac{1}{(1+r)^{n+1}}g^+(S_{n+1})\right]$$

for the discounted intrinsic value process $\frac{1}{(1+r)^n}g^+(S_n)$. Because this process is a submartingale, Theorem 4.3.3 implies that for every stopping time $\tau \in \mathcal{S}_0$

$$\widetilde{\mathbb{E}}\left[\frac{1}{(1+r)^{N\wedge\tau}}g^+(S_{N\wedge\tau})\right] \le \widetilde{\mathbb{E}}\left[\frac{1}{(1+r)^N}g^+(S_N)\right] = V_0^E. \tag{4.5.7}$$

If $\tau \le N$, then

$$\mathbb{I}_{\{\tau\le N\}}\frac{1}{(1+r)^\tau}g(S_\tau) = \frac{1}{(1+r)^{N\wedge\tau}}g(S_{N\wedge\tau}) \le \frac{1}{(1+r)^{N\wedge\tau}}g^+(S_{N\wedge\tau}),$$

and if $\tau = \infty$, then

$$\mathbb{I}_{\{\tau\le N\}}\frac{1}{(1+r)^\tau}g(S_\tau) = 0 \le \frac{1}{(1+r)^{N\wedge\tau}}g^+(S_{N\wedge\tau}).$$

In either case, we have the same result and so

$$\widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau\le N\}}\frac{1}{(1+r)^\tau}g(S_\tau)\right] \le \widetilde{\mathbb{E}}\left[\frac{1}{(1+r)^{N\wedge\tau}}g^+(S_{N\wedge\tau})\right] \le V_0^E,$$

where we have used (4.5.7) for the last step. Since this last inequality holds for every $\tau \in \mathcal{S}_0$, we must have

$$V_0^A = \max_{\tau\in\mathcal{S}_0}\widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau\le N\}}\frac{1}{(1+r)^\tau}g(S_\tau)\right] \le V_0^E. \qquad \square$$

Theorem 4.5.1 shows that the early exercise feature of the American call contributes nothing to its value. An examination of the proof of the theorem indicates that this is because the discounted intrinsic value of the call is a submartingale (i.e., has a tendency to rise) under the risk-neutral probabilities. The discounted intrinsic value of an American put is not a submartingale. If $g^+(s) = (K-s)^+$, then the Jensen inequality (4.5.5) still holds but (4.5.6) does not. Jensen's inequality says that the convex payoff of the put imparts to the discounted intrinsic value of the put a tendency to rise over time, but this may be overcome by a second effect. Because the owner of the put receives K upon exercise, she may exercise early in order to prevent the value of this payment from being discounted away. For low stock prices, this second effect becomes more important than the convexity, and early exercise becomes optimal. For the call, the owner pays K and prefers that the value of this payment be discounted away before exercise. This reinforces the convexity effect and makes early exercise undesirable.

4.6 Summary

Unlike a European derivative security, which can only be exercised at one time, the so-called *expiration date*, an American derivative security entitles its owner to exercise at any time prior to or at the expiration date. One consequence of this is that the value of an American derivative security is always at least as great as the payoff its owner would receive from immediate exercise, the so-called *intrinsic value*. Unlike a European derivative security, whose discounted value is a martingale under the risk-neutral measure, the discounted value of an American derivative security is a supermartingale under the risk-neutral measure. It has a tendency to go down at exactly those moments when it should be exercised. In fact, the value process of an American derivative security is the smallest nonnegative process dominating the intrinsic value and that is a supermartingale under the risk-neutral measure, when discounted. This is the content of Theorem 4.4.2. That theorem leads to the American derivative security pricing algorithm of Theorem 4.4.3:

$$V_N(\omega_1 \dots \omega_N) = \max\{G_N(\omega_1 \dots \omega_N), 0\}, \tag{4.4.10}$$

$$V_n(\omega_1 \dots \omega_n) = \max\Bigg\{G_n(\omega_1 \dots \omega_n), \frac{1}{1+r}\Big[\tilde{p}V_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}V_{n+1}(\omega_1 \dots \omega_n T)\Big]\Bigg\}, \tag{4.4.11}$$

for $n = N-1, \dots, 0$. Here $G_n(\omega_1 \dots \omega_n)$ is the intrinsic value of the security at time n if the first n coin tosses result in $\omega_1 \dots \omega_n$.

The supermartingale property for the discounted American derivative security value process permits an agent holding a short position in the security to construct a hedge. This hedge is constructed by the same formula used for a European derivative security:

$$\Delta_n(\omega_1 \dots \omega_n) = \frac{V_{n+1}(\omega_1 \dots \omega_n H) - V_{n+1}(\omega_1 \dots \omega_n T)}{S_{n+1}(\omega_1 \dots \omega_n H) - S_{n+1}(\omega_1 \dots \omega_n T)}. \tag{4.4.14}$$

In those periods n in which the discounted value process has a strictly downward trend, the short position hedger can consume

$$C_n(\omega_1 \dots \omega_n) = V_n(\omega_1 \dots \omega_n) - \frac{1}{1+r}\Big[\tilde{p}V_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}V_{n+1}(\omega_1 \dots \omega_n T)\Big] \tag{4.4.15}$$

and still maintain the hedge. This is the content of Theorem 4.4.4.

The owner of an American derivative security should exercise the first time the value of the security agrees with its intrinsic value. This exercise rule results in a payment whose discounted risk-neutral value at time zero agrees with the price of the American derivative security at time zero (i.e., this rule permits the owner to capture the full value of the American derivative security). This is the content of Theorem 4.4.5.

The proofs of the claims above use the idea of a stopping time, a random time that makes the decision to stop (exercise) without looking ahead; see Definition 4.3.1. The exercise strategy of the owner of an American derivative security should be a stopping time (i.e., it may depend on past stock price movements but must make the decision to exercise without looking at future price movements). Once a stopping time is chosen, one can compute the risk-neutral expected discounted payoff of the derivative security when that stopping time is used. The value of an American derivative security is the maximum of these risk-neutral expected discounted payoffs over all stopping times. This is Definition 4.4.1.

Martingales, supermartingale, and submartingales evaluated at stopping times rather than nonrandom times have the same trends as if they were evaluated at random times. In particular, if X_n, $n = 0, 1, \dots, N$ is a submartingale under $\widetilde{\mathbb{P}}$ and τ is a stopping time, then $X_{n\wedge\tau}$, $n = 0, 1, \dots, N$ is also a submartingale under $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{E}}X_{n\wedge\tau} \le \widetilde{\mathbb{E}}X_n$. Results of this type are called *optional sampling*. Using optional sampling, one can show that the value of an American call on a stock paying no dividends is the same as the value of a European call on the same stock (i.e., the early exercise option in the American call has zero value). This is the content of Theorem 4.5.1 .

4.7 Notes

A rigorous analysis of American derivative securities based on stopping times was initiated by Bensoussan [2] and continued by Karatzas [26]. A comprehensive treatment appears in Karatzas and Shreve [28]. All of this material treats continuous-time models; it has been specialized to the binomial model in this text.

4.8 Exercises

Exercise 4.1. In the three-period model of Figure 1.2.2 of Chapter 1, let the interest rate be $r = \frac{1}{4}$ so the risk-neutral probabilities are $\tilde{p} = \tilde{q} = \frac{1}{2}$.

(i) Determine the price at time zero, denoted V_0^P, of the American put that expires at time three and has intrinsic value $g_P(s) = (4-s)^+$.
(ii) Determine the price at time zero, denoted V_0^C, of the American call that expires at time three and has intrinsic value $g_C(s) = (s-4)^+$.
(iii) Determine the price at time zero, denoted V_0^S, of the American straddle that expires at time three and has intrinsic value $g_S(s) = g_P(s) + g_C(s)$.
(iv) Explain why $V_0^S < V_0^P + V_0^C$.

Exercise 4.2. In Example 4.2.1, we computed the time-zero value of the American put with strike price 5 to be 1.36. Consider an agent who borrows 1.36 at time zero and buys the put. Explain how this agent can generate sufficient funds to pay off his loan (which grows by 25% each period) by trading in the stock and money markets and optimally exercising the put.

Exercise 4.3. In the three-period model of Figure 1.2.2 of Chapter 1, let the interest rate be $r = \frac{1}{4}$ so the risk-neutral probabilities are $\tilde{p} = \tilde{q} = \frac{1}{2}$. Find the time-zero price and optimal exercise policy (optimal stopping time) for the path-dependent American derivative security whose intrinsic value at each time n, $n = 0, 1, 2, 3$, is $\left(4 - \frac{1}{n+1}\sum_{j=0}^{n} S_j\right)^+$. This intrinsic value is a put on the average stock price between time zero and time n.

Exercise 4.4. Consider the American put of Example 4.2.1, which has strike price 5. Suppose at time zero we sell this put to a purchaser who has inside information about the stock movements and uses the exercise rule ρ of (4.3.2). In particular, if the first toss is going to result in H, the owner of the put exercises at time zero, when the put has intrinsic value 1. If the first toss results in T and the second toss is going to result in H, the owner exercises at time one, when the put has intrinsic value 3. If the first two tosses result in TT, the owner exercises at time two, when the intrinsic value is 4. In summary, the owner of the put has the payoff random variable

$$Y(HH) = 1,\ Y(HT) = 1,\ Y(TH) = 3,\ Y(TT) = 4. \tag{4.8.1}$$

The risk-neutral expected value of this payoff, discounted from the time of payment back to zero, is

$$\widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{\rho} Y\right] = \frac{1}{4}\left[1 + 1 + \frac{4}{5}\cdot 3 + \frac{16}{25}\cdot 4\right] = 1.74. \tag{4.8.2}$$

The time-zero price of the put computed in Example 4.2.1 is only 1.36. Do we need to charge the insider more than this amount if we are going to successfully hedge our short position after selling the put to her? Explain why or why not.

Exercise 4.5. In equation (4.4.5), the maximum is computed over all stopping times in $\mathcal{S}_0$. List all the stopping times in $\mathcal{S}_0$ (there are 26), and from among them, list the stopping times that never exercise when the option is out of the money (there are 11). For each stopping time τ in the latter set, compute $\mathbb{E}\left[\mathbb{I}_{\{\tau\le 2\}}\left(\frac{4}{5}\right)^{\tau} G_\tau\right]$. Verify that the largest value for this quantity is given by the stopping time of (4.4.6), the one that makes this quantity equal to the 1.36 computed in (4.4.7).

Exercise 4.6 (Estimating American put prices). For each n, where $n = 0, 1, \ldots, N$, let G_n be a random variable depending on the first n coin tosses. The time-zero value of a derivative security that can be exercised at any time $n \le N$ for payoff G_n but *must be exercised at time N if it has not been exercised before that time* is

$$V_0 = \max_{\tau\in\mathcal{S}_0,\tau\le N} \widetilde{\mathbb{E}}\left[\frac{1}{(1+r)^{\tau}} G_\tau\right]. \tag{4.8.3}$$

In contrast to equation (4.4.1) in the Definition 4.4.1 for American derivative securities, here we consider only stopping times that take one of the values $0, 1, \ldots, N$ and not the value ∞.

(i) Consider $G_n = K - S_n$, the derivative security that permits its owner to sell one share of stock for payment K at any time up to and including N, but if the owner does not sell by time N, then she must do so at time N. Show that the optimal exercise policy is to sell the stock at time zero and that the value of this derivative security is $K - S_0$.

(ii) Explain why a portfolio that holds the derivative security in (i) and a European call with strike K and expiration time N is at least as valuable as an American put struck at K with expiration time N. Denote the time-zero value of the European call by V_0^{EC} and the time-zero value of the American put by V_0^{AP}. Conclude that the upper bound

$$V_0^{AP} \le K - S_0 + V_0^{EC} \tag{4.8.4}$$

on V_0^{AP} holds.

(iii) Use put–call parity (Exercise 2.11 of Chapter 2) to derive the lower bound on V_0^{AP}:

$$\frac{K}{(1+r)^N} - S_0 + V_0^{EC} \le V_0^{AP}. \tag{4.8.5}$$

Exercise 4.7. For the class of derivative securities described in Exercise 4.6 whose time-zero price is given by (4.8.3), let $G_n = S_n - K$. This derivative security permits its owner to buy one share of stock in exchange for a payment of K at any time up to the expiration time N. If the purchase has not been made at time N, it must be made then. Determine the time-zero value and optimal exercise policy for this derivative security.

5

Random Walk

5.1 Introduction

In this section, we consider a symmetric random walk, which is the discrete-time version of Brownian motion, introduced in Chapter 3 of Volume II. We derive several properties of a random walk, and shall ultimately see that Brownian motion has similar properties. In particular, in this chapter we consider *first passage times* and the *reflection principle* for a symmetric random walk. For Brownian motion, these concepts are used in the computation of the price of a variety of exotic options.

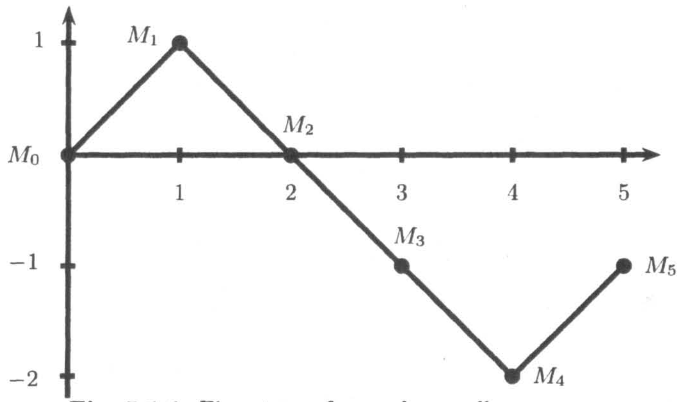

Fig. 5.1.1. Five steps of a random walk

To construct a symmetric random walk, we repeatedly toss a fair coin (p, the probability of H on each toss, and $q = 1 - p$, the probability of T on each toss, are both equal to $\frac{1}{2}$). We denote the successive outcomes of the tosses by $\omega_1\omega_2\omega_3\ldots$. Let

$$X_j = \begin{cases} 1, & \text{if } \omega_j = H, \\ -1, & \text{if } \omega_j = T, \end{cases} \tag{5.1.1}$$

and define $M_0 = 0$,

$$M_n = \sum_{j=1}^{n} X_j, \; n = 1, 2, \ldots. \tag{5.1.2}$$

The process M_n, $n = 0, 1, 2, \ldots$ is a *symmetric random walk*. With each toss, it either steps up one unit or down one unit, and each of the two possibilities is equally likely. If $p \neq \frac{1}{2}$, we would still have a random walk, but it would be *asymmetric*. In other words, the symmetric and asymmetric random walks have the same set of possible paths; they differ only in the assignment of probabilities to these paths.

The symmetric random walk is both a martingale and a Markov process.

5.2 First Passage Times

The symmetric random walk of Section 5.1 starts at 0 at time zero. Fix an integer m, and let τ_m denote the first time the random walk reaches level m; i.e.,

$$\tau_m = \min\{n; M_n = m\}. \tag{5.2.1}$$

If the random walk never reaches the level m, we define τ_m to be infinity.

The random variable τ_m is a stopping time, called the *first passage time* of the random walk to level m. We shall determine its distribution. We shall see that τ_m is finite with probability 1 (i.e., with probability 1, the random walk eventually reaches the level m), but $\mathbb{E}\tau_m = \infty$. Once we have determined the distribution of τ_m for a symmetric random walk, we shall see how to modify the formulas to obtain information about the distribution of τ_m when the walk is asymmetric.

Our study of the distribution of τ_m for a symmetric random walk uses the martingale (5.2.2) below, which is discussed in Exercise 2.4(ii) at the end of Chapter 2.

Lemma 5.2.1. *Let M_n be a symmetric random walk. Fix a number σ and define the process*

$$S_n = e^{\sigma M_n}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^n. \tag{5.2.2}$$

Then S_n, $n = 0, 1, 2, \ldots$ is a martingale.

PROOF: In the notation of (5.1.1) and (5.1.2), we have

$$S_{n+1} = S_n\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)e^{\sigma X_{n+1}}.$$

We take out what is known (Theorem 2.3.2(ii)) and use independence (Theorem 2.3.2(iv)) to write

$$\begin{aligned}\mathbb{E}_n S_{n+1} &= S_n\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)\mathbb{E}e^{\sigma X_{n+1}}\\ &= S_n\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)\left(\frac{1}{2}e^{\sigma}+\frac{1}{2}e^{-\sigma}\right)\\ &= S_n.\end{aligned}$$

This shows that the process S_n, $n = 0, 1, 2, \ldots$ is a martingale. □

Because a martingale stopped at a stopping time is still a martingale (Theorem 4.3.2), the process $S_{n\wedge\tau_m}$ is a martingale and hence has constant expectation, i.e.,

$$1 = S_0 = \mathbb{E}S_{n\wedge\tau_m} = \mathbb{E}\left[e^{\sigma M_{n\wedge\tau_m}}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{n\wedge\tau_m}\right] \text{ for all } n \geq 0. \tag{5.2.3}$$

We would like to let $n \to \infty$ in (5.2.3). In order to do that, we must determine the limit as $n \to \infty$ of $e^{\sigma M_{n\wedge\tau_m}}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{n\wedge\tau_m}$. We treat the two factors separately.

We observe first that $\cosh(\sigma) = \frac{e^{\sigma}+e^{-\sigma}}{2}$ attains its minimum at $\sigma = 0$, where it takes the value 1. Hence, $\cosh(\sigma) > 1$ for all $\sigma > 0$, so

$$0 < \frac{2}{e^{\sigma}+e^{-\sigma}} < 1 \text{ for all } \sigma > 0. \tag{5.2.4}$$

We now fix $\sigma > 0$ and conclude from (5.2.4) that

$$\lim_{n\to\infty}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{n\wedge\tau_m} = \begin{cases}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{\tau_m}, & \text{if } \tau_m < \infty,\\ 0, & \text{if } \tau_m = \infty.\end{cases} \tag{5.2.5}$$

We may write the right-hand side of (5.2.5) as $\mathbb{I}_{\{\tau_m<\infty\}}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{\tau_m}$, which captures both cases.

To study the other factor, $e^{\sigma M_{n\wedge\tau_m}}$, we assume that $m > 0$ and note that $M_{n\wedge\tau_m} \leq m$ because we stop this martingale when it reaches the level m. Hence, regardless of whether τ_m is finite or infinite, we have

$$0 \leq e^{\sigma M_{n\wedge\tau_m}} \leq e^{\sigma m}. \tag{5.2.6}$$

We also have

$$\lim_{n\to\infty} e^{\sigma M_{n\wedge\tau_m}} = e^{\sigma M_{\tau_m}} = e^{\sigma m} \text{ if } \tau_m < \infty. \tag{5.2.7}$$

Taking the product of (5.2.5) and (5.2.7), we see that

$$\lim_{n\to\infty} e^{\sigma M_{n\wedge\tau_m}}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{n\wedge\tau_m} = e^{\sigma m}\left(\frac{2}{e^{\sigma}+e^{-\sigma}}\right)^{\tau_m} \text{ if } \tau_m < \infty. \tag{5.2.8}$$

We don't know that $e^{\sigma M_{n\wedge\tau_m}}$ has a limit as $n\to\infty$ along the paths for which $\tau_m=\infty$, but that does not matter because this term is bounded (see (5.2.6)) and $\left(\frac{2}{e^\sigma+e^{-\sigma}}\right)^{n\wedge\tau_m}$ has limit zero. Hence,

$$\lim_{n\to\infty} e^{\sigma M_{n\wedge\tau_m}}\left(\frac{2}{e^\sigma+e^{-\sigma}}\right)^{n\wedge\tau_m}=0 \text{ if } \tau_m=\infty. \tag{5.2.9}$$

It follows from (5.2.8) and (5.2.9) that

$$\lim_{n\to\infty} e^{\sigma M_{n\wedge\tau_m}}\left(\frac{2}{e^\sigma+e^{-\sigma}}\right)^{n\wedge\tau_m}=\mathbb{I}_{\{\tau_m<\infty\}}e^{\sigma m}\left(\frac{2}{e^\sigma+e^{-\sigma}}\right)^{\tau_m}. \tag{5.2.10}$$

We may now take the limit in (5.2.3) to obtain[1]

$$\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}}e^{\sigma m}\left(\frac{2}{e^\sigma+e^{-\sigma}}\right)^{\tau_m}\right]=1. \tag{5.2.11}$$

Equation (5.2.11) was derived under the assumption that σ is strictly positive, the assumption we used to derive (5.2.5). Equation (5.2.11) thus holds for all strictly positive σ. We may not set $\sigma=0$ in (5.2.11), but we may compute the limit of both sides[2] as $\sigma\downarrow 0$, and this yields $\mathbb{E}\mathbb{I}_{\{\tau_m<\infty\}}=1$, i.e.,

$$\mathbb{P}\{\tau_m<\infty\}=1. \tag{5.2.12}$$

This tells us that, with probability 1, the symmetric random walk reaches the level m. Nonetheless, there are paths of the symmetric random walk that never reach level m. For example, any path in which the tails obtained at any time always outnumber the heads (such a path might begin with $TTHTTHTHHT...$) will begin at zero and take only strictly negative values thereafter. Such a path never reaches the positive level m. Equation (5.2.12) asserts that although there are infinitely many such paths (in fact, uncountably infinitely many), taken all together, the set of these paths has zero probability. We have assumed for this discussion that m is strictly positive (we used this assumption to derive (5.2.6)), but because of the symmetry of the random walk, the conclusion (5.2.12) also holds if m is strictly negative.

When some event happens with probability 1, we say the event happens *almost surely*. We have proved the following theorem.

Theorem 5.2.2. *Let m be an arbitrary nonzero integer. The symmetric random walk reaches the level m almost surely; i.e., the first passage time τ_m to level m is finite almost surely.*

[1] It is not always possible to conclude from convergence of a sequence of random variables that the expectations converge in the same way. In this particular case, we may take this step because the random variables appearing in (5.2.10) are bounded between two constants, 0 and $e^{\sigma m}$. The applicable theorem is the Dominated Convergence Theorem 1.4.9 of Volume II, Chapter 1.

[2] The computation of the limit of the left-hand side of (5.2.11) requires another application of the Dominated Convergence Theorem 1.4.9 of Volume II, Chapter 1.

In order to determine the distribution of the first passage time τ_m, we examine its moment-generating function $\varphi_{\tau_m}(u)=\mathbb{E}e^{u\tau_m}$. For all $x\ge 0$, $e^x=1+x+\frac{1}{2}x^2+\cdots\ge x$. Hence, for positive u, we have $e^{u\tau_m}\ge u\tau_m$ and so $\varphi_{\tau_m}(u)=\mathbb{E}e^{u\tau_m}\ge u\mathbb{E}\tau_m$. We shall see in Corollary 5.2.4 below that $\mathbb{E}\tau_m=\infty$, which implies that $\varphi_{\tau_m}(u)=\infty$ for $u>0$. For $u=0$, we have $\varphi_{\tau_m}(u)=1$. Therefore, the moment-generating function $\varphi_{\tau_m}(u)$ is interesting only for $u<0$. For $u<0$, we set $\alpha=e^u$ so that $0<\alpha<1$ and $\varphi_{\tau_m}(u)=\mathbb{E}\alpha^{\tau_m}$.

Theorem 5.2.3. *Let m be a nonzero integer. The first passage time τ_m for the symmetric random walk satisfies*

$$\mathbb{E}\alpha^{\tau_m}=\left(\frac{1-\sqrt{1-\alpha^2}}{\alpha}\right)^{|m|} \quad \textit{for all } \alpha\in(0,1). \tag{5.2.13}$$

PROOF: For the symmetric random walk, τ_m and τ_{-m} have the same distribution, so it is enough to prove the theorem for the case that m is a positive integer. We take m to be a positive integer. Because $\mathbb{P}\{\tau_m<\infty\}=1$, we may simplify (5.2.11) to

$$\mathbb{E}\left[e^{\sigma m}\left(\frac{2}{e^\sigma+e^{-\sigma}}\right)^{\tau_m}\right]=1. \tag{5.2.14}$$

This equation holds for all strictly positive σ.

To obtain (5.2.13) from (5.2.14), we let $\alpha\in(0,1)$ be given and solve for $\sigma>0$, which satisfies

$$\alpha=\frac{2}{e^\sigma+e^{-\sigma}}. \tag{5.2.15}$$

This is equivalent to

$$\alpha e^\sigma+\alpha e^{-\sigma}-2=0,$$

which is in turn equivalent to

$$\alpha\left(e^{-\sigma}\right)^2-2e^{-\sigma}+\alpha=0.$$

This last equation is a quadratic equation in the unknown $e^{-\sigma}$, and the solutions are

$$e^{-\sigma}=\frac{2\pm\sqrt{4-4\alpha^2}}{2\alpha}=\frac{1\pm\sqrt{1-\alpha^2}}{\alpha}.$$

We need to find a strictly positive σ satisfying this equation, and thus we need $e^{-\sigma}$ to be strictly less than one. That suggests we should take the solution for $e^{-\sigma}$ corresponding to the negative sign in the formula above; i.e.,

$$e^{-\sigma}=\frac{1-\sqrt{1-\alpha^2}}{\alpha}. \tag{5.2.16}$$

We verify that (5.2.16) leads to a value of σ that is strictly positive. Because we have chosen α to satisfy $0<\alpha<1$, we have

$$0 < (1-\alpha)^2 < 1 - \alpha < 1 - \alpha^2.$$

Taking positive square roots, we obtain

$$1 - \alpha < \sqrt{1-\alpha^2}.$$

Therefore, $1 - \sqrt{1-\alpha^2} < \alpha$, and dividing through by α, we see that the right-hand side of (5.2.16) is strictly less than 1. This implies that σ in (5.2.16) is strictly positive.

With σ and α related by (5.2.16), we have that α and σ are also related by (5.2.15) and hence may rewrite (5.2.14) as

$$\mathbb{E}\left[\left(\frac{\alpha}{1-\sqrt{1-\alpha^2}}\right)^m \alpha^{\tau_m}\right] = 1.$$

Because $\left(\frac{\alpha}{1-\sqrt{1-\alpha^2}}\right)^m$ is not random, we may take it outside the expectation and divide through by it. Equation (5.2.13) for positive m follows immediately.
□

Corollary 5.2.4. *Under the conditions of Theorem 5.2.3, we have*

$$\mathbb{E}\tau_m = \infty. \tag{5.2.17}$$

PROOF:[3] We first show that $\mathbb{E}\tau_1 = \infty$. To do this, we differentiate both sides of (5.2.13) with respect to α:

$$\begin{aligned}
\mathbb{E}\left[\tau_1 \alpha^{\tau_1 - 1}\right] &= \frac{\partial}{\partial \alpha}\mathbb{E}\alpha^{\tau_1} \\
&= \frac{\partial}{\partial \alpha}\frac{1-\sqrt{1-\alpha^2}}{\alpha} \\
&= \frac{-\frac{1}{2}(1-\alpha^2)^{-\frac{1}{2}}(-2\alpha)\alpha - \left(1-(1-\alpha^2)^{\frac{1}{2}}\right)}{\alpha^2} \\
&= \frac{\alpha^2(1-\alpha^2)^{-\frac{1}{2}} - 1 + (1-\alpha^2)^{\frac{1}{2}}}{\alpha^2} \\
&= \frac{\alpha^2 - \sqrt{1-\alpha^2} + 1 - \alpha^2}{\alpha^2\sqrt{1-\alpha^2}} \\
&= \frac{1-\sqrt{1-\alpha^2}}{\alpha^2\sqrt{1-\alpha^2}}.
\end{aligned}$$

This equation is valid for all $\alpha \in (0,1)$. We may not substitute $\alpha = 1$ into this equation, but we may take the limit of both sides as $\alpha \uparrow 1$, and this gives us $\mathbb{E}\tau_1 = \infty$.

[3] There are two steps in the proof of Corollary 5.2.4 that require the interchange of limit and expectation. The first of these, differentiation of $\mathbb{E}\alpha^{\tau_1}$ with respect to α, can be justified by an argument like that in Exercise 8 in Volume II, Chapter 1. The second, where we let $\alpha \uparrow 1$, is an application of the Monotone Convergence Theorem 1.4.5 of Volume II, Chapter 1.

For $m \geq 1$, we have $\tau_m \geq \tau_1$ and hence $\mathbb{E}\tau_m \geq \mathbb{E}\tau_1 = \infty$. For strictly negative integers m, the symmetry of the random walk now implies $\mathbb{E}\tau_m = \infty$.
□

From the formula (5.2.13), it is possible to compute explicitly the distribution of the random variable τ_1. We have the special case of formula (5.2.13):

$$\mathbb{E}\alpha^{\tau_1} = \frac{1-\sqrt{1-\alpha^2}}{\alpha} \quad \text{for all } \alpha \in (0,1). \tag{5.2.18}$$

Because the random walk can reach level 1 only on an odd-numbered step, the left-hand side of (5.2.18) may be rewritten as

$$\mathbb{E}\alpha^{\tau_1} = \sum_{j=1}^{\infty} \alpha^{2j-1}\mathbb{P}\{\tau_1 = 2j-1\} \tag{5.2.19}$$

We work out a power series expansion for the right-hand side of (5.2.18). Define $f(x) = 1 - \sqrt{1-x}$ so that

$$\begin{aligned}
f'(x) &= \frac{1}{2}(1-x)^{-\frac{1}{2}}, \\
f''(x) &= \frac{1}{4}(1-x)^{-\frac{3}{2}}, \\
f'''(x) &= \frac{3}{8}(1-x)^{-\frac{5}{2}},
\end{aligned}$$

and, in general, the jth-order derivative of f is

$$f^{(j)}(x) = \frac{1\cdot 3\cdots(2j-3)}{2^j}(1-x)^{-\frac{2j-1}{2}}, \quad j = 1,2,3,\ldots.$$

Evaluating at 0, we obtain

$$f(0) = 0,\ f'(0) = \frac{1}{2},\ f''(0) = \frac{1}{4},\ f'''(0) = \frac{3}{8}$$

and, in general,

$$\begin{aligned}
f^{(j)}(0) &= \frac{1\cdot 3\ldots(2j-3)}{2^j} \\
&= \frac{1\cdot 3\ldots(2j-3)}{2^j}\cdot\frac{2\cdot 4\cdots(2j-2)}{2^{j-1}(j-1)!} \\
&= \left(\frac{1}{2}\right)^{2j-1}\frac{(2j-2)!}{(j-1)!}, \quad j = 1,2,3,\ldots.
\end{aligned} \tag{5.2.20}$$

(We use here the definition $0! = 1$.) The Taylor series expansion of $f(x)$ is

$$f(x) = 1 - \sqrt{1-x} = \sum_{j=0}^{\infty}\frac{1}{j!}f^{(j)}(0)x^j = \sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{2j-1}\frac{(2j-2)!}{j!(j-1)!}x^j.$$

Therefore,

$$\frac{1-\sqrt{1-\alpha^2}}{\alpha} = \frac{f(\alpha^2)}{\alpha} = \sum_{j=1}^{\infty}\left(\frac{\alpha}{2}\right)^{2j-1}\frac{(2j-2)!}{j!(j-1)!}. \tag{5.2.21}$$

We have thus worked out the power series (5.2.19) and (5.2.21) for the two sides of (5.2.18). Equating them, we obtain

$$\sum_{j=1}^{\infty}\alpha^{2j-1}\mathbb{P}\{\tau_1 = 2j-1\} = \sum_{j=1}^{\infty}\left(\frac{\alpha}{2}\right)^{2j-1}\frac{(2j-2)!}{j!(j-1)!} \quad \text{for all } \alpha \in (0,1).$$

The only way these two power series can be equal for all $\alpha \in (0,1)$ is for their coefficients to be equal term-by-term (i.e., the term multiplying α^{2j-1} must be the same in both series). This gives us the formula

$$\mathbb{P}\{\tau_1 = 2j-1\} = \frac{(2j-2)!}{j!(j-1)!}\cdot\left(\frac{1}{2}\right)^{2j-1}, \quad j = 1,2,\dots. \tag{5.2.22}$$

We verify (5.2.22) for the first few values of j. For $j=1$, we have

$$\mathbb{P}\{\tau_1 = 1\} = \frac{0!}{1!0!}\cdot\frac{1}{2} = \frac{1}{2}.$$

The only way τ_1 can be 1 is for the first coin toss to result in H, and the probability of this for a symmetric random walk is $\frac{1}{2}$. For $j=2$, we have

$$\mathbb{P}\{\tau_1 = 3\} = \frac{2!}{2!1!}\cdot\left(\frac{1}{2}\right)^3 = \left(\frac{1}{2}\right)^3.$$

The only way τ_1 can be 3 is for the first three coin tosses to result in THH, and the probability of this is $\left(\frac{1}{2}\right)^3$. For $j=3$, we have

$$\mathbb{P}\{\tau_1 = 5\} = \frac{4!}{3!2!}\cdot\left(\frac{1}{2}\right)^5 = 2\cdot\left(\frac{1}{2}\right)^5.$$

There are two ways τ_1 can be 5; the first five tosses could be either $THTHH$ or $TTHHH$, and the probability of each of these outcomes is $\left(\frac{1}{2}\right)^5$.

From these examples, we see how to interpret the two factors appearing on the right-hand side of (5.2.22). The term $\frac{(2j-2)!}{j!(j-1)!}$ counts the number of paths with $2j-1$ steps that first reach level 1 on the $(2j-1)$st step so that $\tau_1 = 2j-1$. The term $\left(\frac{1}{2}\right)^{2j-1}$ is the probability of each of these paths. Suppose now that the random walk is not symmetric, but has probability p for H and probability $q = 1-p$ for T. The number of paths that first reach level 1 on the $(2j-1)$st step is unaffected. However, since each of these paths must have j up steps and $j-1$ down steps, the probability of such a path is now $p^j q^{j-1}$. This observation leads to the following theorem.

Theorem 5.2.5. *Let τ_1 be the first passage time to level 1 of a random walk that has probability p for an up step and probability $q = 1-p$ for a down step. Then*

$$\mathbb{P}\{\tau_1 = 2j-1\} = \frac{(2j-2)!}{j!(j-1)!}p^j q^{j-1}, \quad j = 1,2,\dots. \tag{5.2.23}$$

5.3 Reflection Principle

In this section, we give a second proof of Theorem 5.2.5, based on the *reflection principle*. We shall use this same idea in our study of Brownian motion. As in Section 5.2, we first consider a symmetric random walk, obtaining formula (5.2.22). The remainder of the argument to obtain Theorem 5.2.5 is the same as in Section 5.2.

Suppose we toss a coin an odd number $(2j-1)$ of times. Some of the paths of the random walk will reach the level 1 in the first $2j-1$ steps and others will not. In the case of three tosses, there are eight possible paths and five of these reach the level 1 (see Figure 5.3.1). Consider a path that reaches the level 1 at some time $\tau_1 \le 2j-1$. From that moment on, we can create a "reflected" path, which steps up each time the original path steps down and steps down each time the original path steps up. If the original path ends above 1 at the final time $2j-1$, the reflected path ends below 1, and vice versa. If the original path ends at 1, the reflected path does also.

To count the number of paths that reach the level 1 by time $2j-1$, we can count the number of paths that exceed 1 at time $2j-1$, the number of paths that are at 1 at time $2j-1$, and the number of reflected paths that exceed 1 at time $2j-1$. The reflected paths that exceed 1 correspond to paths that reached 1 at some time prior to $2j-1$ but are below 1 at time $2j-1$. In Figure 5.3.1, there is only one path that exceeds 1 at time three: HHH. There are three paths that are at 1 at time three: HHT, HTH, and THH. There is one path whose reflected path exceeds 1 at time three: HTT. These account for all five paths that reach level 1 by time three.

There are as many reflected paths that exceed 1 at time $2j-1$ as there are original paths that exceed 1 at time $2j-1$. In Figure 5.3.1, there is one of each of these. Thus, to count the number of paths that reach level 1 by time $2j-1$, we can count the paths that are at 1 at time $2j-1$ and then add on *twice* the number of paths that exceed 1 at time $2j-1$. In other words, for the symmetric random walk,

$$\mathbb{P}\{\tau_1 \le 2j-1\} = \mathbb{P}\{M_{2j-1} = 1\} + 2\mathbb{P}\{M_{2j-1} \ge 3\}.$$

But for the symmetric random walk, $\mathbb{P}\{M_{2j-1} \ge 3\} = \mathbb{P}\{M_{2j-1} \le -3\}$, so

$$\begin{aligned}\mathbb{P}\{\tau_1 \le 2j-1\} &= \mathbb{P}\{M_{2j-1} = 1\} + \mathbb{P}\{M_{2j-1} \ge 3\} + \mathbb{P}\{M_{2j-1} \le -3\}\\ &= 1 - \mathbb{P}\{M_{2j-1} = -1\}.\end{aligned}$$

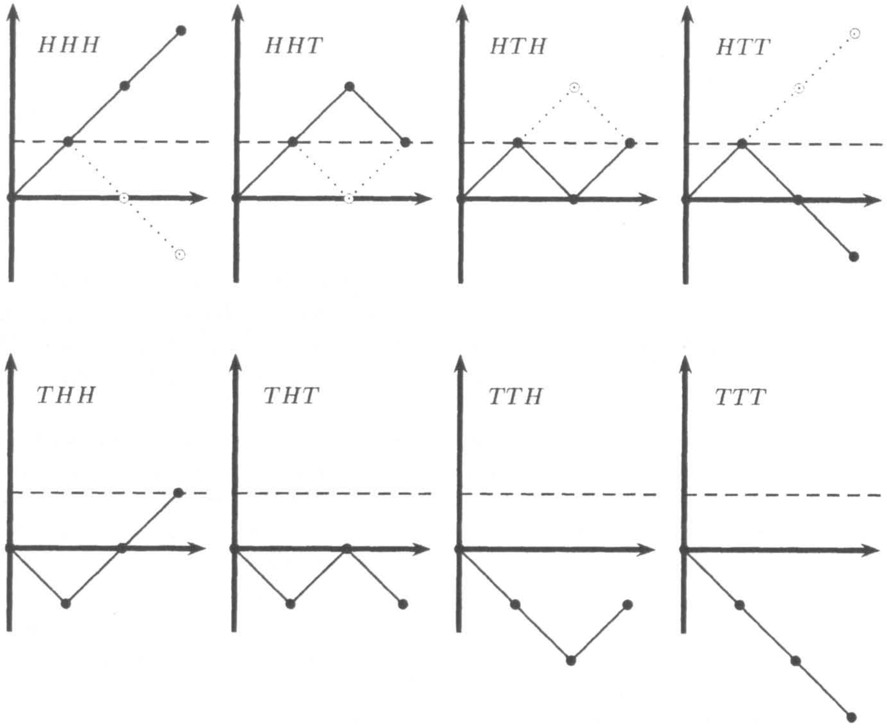

Fig. 5.3.1. Eight paths and four reflected paths.

In order for M_{2j-1} to be -1, in the first $2j-1$ tosses there must be $j-1$ heads and j tails. There are

$$\binom{2j-1}{j} = \frac{(2j-1)!}{j!(j-1)!}$$

paths that have this property, and each such path has probability $\left(\frac{1}{2}\right)^{2j-1}$. Hence,

$$\mathbb{P}\{M_{2j-1} = -1\} = \left(\frac{1}{2}\right)^{2j-1} \frac{(2j-1)!}{j!(j-1)!}.$$

Similarly,

$$\mathbb{P}\{M_{2j-3} = -1\} = \left(\frac{1}{2}\right)^{2j-3} \frac{(2j-3)!}{(j-1)!(j-2)!}.$$

It follows that, for $j \geq 2$,

$$\begin{aligned}
\mathbb{P}\{\tau_1 = 2j-1\} &= \mathbb{P}\{\tau_1 \leq 2j-1\} - \mathbb{P}\{\tau_1 \leq 2j-3\} \\
&= \mathbb{P}\{M_{2j-3} = -1\} - \mathbb{P}\{M_{2j-1} = -1\} \\
&= \left(\frac{1}{2}\right)^{2j-3} \frac{(2j-3)!}{(j-1)!(j-2)!} - \left(\frac{1}{2}\right)^{2j-1} \frac{(2j-1)!}{j!(j-1)!} \\
&= \left(\frac{1}{2}\right)^{2j-1} \frac{(2j-3)!}{j!(j-1)!}\left[4j(j-1) - (2j-1)(2j-2)\right] \\
&= \left(\frac{1}{2}\right)^{2j-1} \frac{(2j-3)!}{j!(j-1)!}\left[2j(2j-2) - (2j-1)(2j-2)\right] \\
&= \left(\frac{1}{2}\right)^{2j-1} \frac{(2j-3)!}{j!(j-1)!}(2j-2) \\
&= \left(\frac{1}{2}\right)^{2j-1} \frac{(2j-2)!}{j!(j-1)!}.
\end{aligned}$$

We have again obtained (5.2.22).

5.4 Perpetual American Put: An Example

In this section, we work out the pricing and hedging for a particular example of a *perpetual American put.* "Perpetual" refers to the fact that this put has no expiration date. This is not a traded instrument but rather a mathematical concept that serves as a bridge between the discrete-time American pricing and hedging discussion of Chapter 4 and the continuous-time analysis for American derivative securities in Chapter 8 of Volume II.

We consider a binomial model with up factor $u = 2$, down factor $d = \frac{1}{2}$, and interest rate $r = \frac{1}{4}$. For this model, the risk-neutral probabilities (see (1.1.8) of Chapter 1) are $\tilde{p} = \tilde{q} = \frac{1}{2}$. The stock price at time n is

$$S_n = S_0 \cdot 2^{M_n}, \tag{5.4.1}$$

where M_n is the random walk of (5.1.2). Under the risk-neutral probabilities, M_n is symmetric.

Consider an American put with strike price $K = 4$ and no expiration date. At any time n, the owner of this put can exercise it, selling for \$4 a share of stock worth \$$S_n$. We are interested in the value of this put as a function of the underlying stock price. Because there is no expiration, it is reasonable to expect the value of the put to depend only on the stock price, not on time. Similarly, it is reasonable to expect the optimal exercise policy to depend only on the stock price, not on time.

For the moment, let us suppose that $S_0 = 4$. Here are some possible exercise policies:

Policy 0: Exercise immediately. This corresponds to the stopping time τ_0, the first time the random walk M_n reaches the level 0, which is $\tau_0 = 0$. The associated value of this exercise policy is $V^{(\tau_0)} = 0$.

Policy 1: Exercise the first time the stock price falls to the level 2. This time is τ_{-1}, the first time the random walk falls to the level -1. We denote by $V^{(\tau_{-1})}$ the value of this exercise policy and compute it below.
Policy 2: Exercise the first time the stock price falls to the level 1. This time is τ_{-2}, the first time the random walk falls to the level -2. We denote by $V^{(\tau_{-2})}$ the value of this exercise policy and compute it below.

Let m be a positive integer. The risk-neutral value of the put if the owner uses the exercise policy τ_{-m}, which exercises the first time the stock price falls to $4 \cdot 2^{-m}$, is

$$V^{(\tau_{-m})} = \widetilde{\mathbb{E}}\left[\left(\frac{1}{1+r}\right)^{\tau_{-m}} (K - S_{\tau_{-m}})\right] = 4(1-2^{-m})\widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{\tau_{-m}}\right]. \quad (5.4.2)$$

This is the risk-neutral expected payoff of the option at the time of exercise, discounted from the exercise time back to time zero. Because M_n is a symmetric random walk under the risk-neutral probabilities, we can compute the right-hand side of (5.4.2) using Theorem 5.2.3. We take α in that theorem to be $\frac{4}{5}$, so that

$$\frac{1-\sqrt{1-\alpha^2}}{\alpha} = \frac{5}{4}\cdot\left(1-\sqrt{1-\left(\frac{4}{5}\right)^2}\right) = \frac{5}{4}\cdot\left(1-\sqrt{\frac{9}{25}}\right) = \frac{5}{4}\cdot\frac{2}{5} = \frac{1}{2}.$$

Theorem (5.2.3) implies that

$$V^{(\tau_{-m})} = 4(1-2^{-m})\left(\frac{1}{2}\right)^m, \quad m = 1, 2, \dots. \quad (5.4.3)$$

In particular,

$$V^{(\tau_{-1})} = 4\left(1-\frac{1}{2}\right)\frac{1}{2} = 1,$$
$$V^{(\tau_{-2})} = 4\left(1-\frac{1}{4}\right)\frac{1}{4} = \frac{3}{4},$$
$$V^{(\tau_{-3})} = 4\left(1-\frac{1}{8}\right)\frac{1}{8} = \frac{7}{16}.$$

Based on the computations above, we guess that the optimal policy is to exercise the first time the stock price falls to 2. This appears to give the option the largest value, at least if $S_0 = 4$. It is reasonable to expect that the optimality of this policy does not depend on the initial stock price. In other words, we expect that, regardless of the initial stock price, one should exercise the option the first time the stock price falls to 2. If the initial stock price is 2 or less, one would then exercise immediately.

To confirm the optimality of the exercise policy described in the previous paragraph, we first determine the value of the option for different initial stock

prices when we use the policy of exercising the first time the stock price is at or below 2. If the initial stock price is $S_0 = 2^j$ for some integer $j \le 1$, then $S_0 \le 2$ and we exercise immediately. The value of the option under this exercise policy is the intrinsic value

$$v(2^j) = 4 - 2^j, \quad j = 1, 0, -1, -2, \dots. \quad (5.4.4)$$

Now suppose the initial stock price is $S_0 = 2^j$ for some integer $j \ge 2$. We use the policy of exercising the first time the stock price reaches 2, which requires the random walk to fall to $-(j-1)$. We obtain the following values for the initial stock prices $S_0 = 4$ and $S_0 = 8$:

$$v(4) = \widetilde{\mathbb{E}}\left[\left(\frac{1}{1+r}\right)^{\tau_{-1}} (K - S_{\tau_{-1}})\right] = 2\widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{\tau_{-1}}\right] = 2\cdot\left(\frac{1}{2}\right)^1 = 1,$$
$$v(8) = \widetilde{\mathbb{E}}\left[\left(\frac{1}{1+r}\right)^{\tau_{-2}} (K - S_{\tau_{-2}})\right] = 2\widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{\tau_{-2}}\right] = 2\cdot\left(\frac{1}{2}\right)^2 = \frac{1}{2}.$$

In general,

$$v(2^j) = \widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{\tau_{-(j-1)}} (4 - S_{\tau_{-(j-1)}})\right] = 2\cdot\left(\frac{1}{2}\right)^{j-1} = \frac{4}{2^j}, \quad j = 2, 3, 4, \dots. \quad (5.4.5)$$

Notice that, when $j = 1$, formula (5.4.5) gives the same value as (5.4.4).

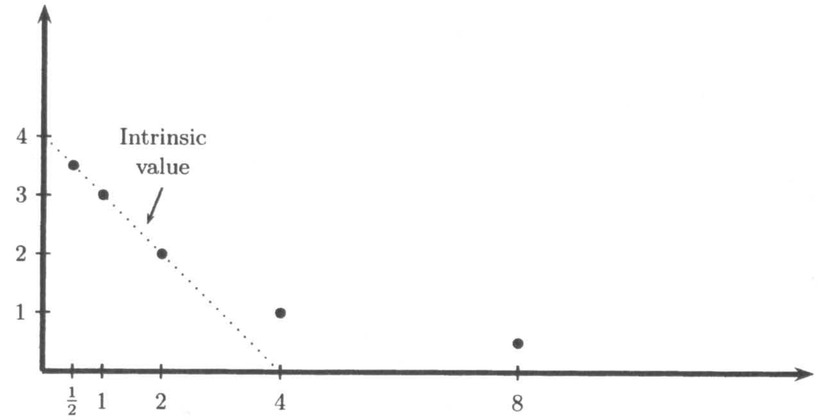

Fig. 5.4.1. Perpetual American put price $v(2^j)$.

We shall only consider stock prices of the form 2^j; if the initial stock price is of this form, then all subsequent stock prices are of this form (but with different integers j). At this point, we have a conjectured optimal exercise policy (exercise as soon as the stock price is 2 or below) and a conjectured option value function

$$v(2^j) = \begin{cases} 4 - 2^j, & \text{if } j \leq 1, \\ \frac{4}{2^j}, & \text{if } j \geq 1. \end{cases} \tag{5.4.6}$$

To verify these conjectures, we need to establish analogues of the three properties in Theorem 4.4.2 of Chapter 4:

(i) $v(S_n) \geq (4 - S_n)^+$, $n = 0, 1, \ldots$;
(ii) the discounted process $\left(\frac{4}{5}\right)^n v(S_n)$ is a supermartingale under the risk-neutral probability measure;
(iii) $v(S_n)$ is the smallest process satisfying (i) and (ii).

Property (ii) guarantees that if we sell the option at time zero for $v(S_0)$, it is possible to take this initial capital and construct a hedge, sometimes consuming, so that the value of our hedging portfolio at each time n is $v(S_n)$. The details of this are given in the proof of Theorem 4.4.4 of Chapter 4. Property (i) guarantees that the value of the hedging portfolio, which is $v(S_n)$ at each time n, is sufficient to pay off the option when it is exercised. Taken together, properties (i) and (ii) guarantee that the seller is satisfied with this option price. Property (iii) guarantees that the buyer is also satisfied.

We now verify that properties (i), (ii), and (iii) are satisfied by the function $v(2^j)$ given by (5.4.6).

Property (i): For $j \leq 1$ and $S_n = 2^j$, (5.4.6) implies immediately that $v(S_n) = 4 - S_n \geq (4 - S_n)^+$. For $j \geq 2$ and $S_n = 2^j$, we have $v(S_n) \geq 0 = (4 - S_n)^+$.

Property (ii): For $S_n = 2^j$ and $j \leq 0$, we have

$$\begin{aligned}
\widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{n+1} v(S_{n+1})\right] &= \frac{1}{2}\left(\frac{4}{5}\right)^{n+1} v(2^{j+1}) + \frac{1}{2}\left(\frac{4}{5}\right)^{n+1} v(2^{j-1}) \\
&= \left(\frac{4}{5}\right)^n \left[\frac{2}{5} v(2^{j+1}) + \frac{2}{5} v(2^{j-1})\right] \\
&= \left(\frac{4}{5}\right)^n \left[\frac{2}{5}(4 - 2^{j+1}) + \frac{2}{5}(4 - 2^{j-1})\right] \\
&= \left(\frac{4}{5}\right)^n \left[\frac{16}{5} - \frac{1}{5}(4+1)2^j\right] \\
&= \left(\frac{4}{5}\right)^n \left[\frac{16}{5} - 2^j\right] \\
&< \left(\frac{4}{5}\right)^n [4 - 2^j] = \left(\frac{4}{5}\right)^n v(S_n).
\end{aligned}$$

We thus see that in the "exercise region" of stock prices 1 and below, the discounted option value is a strict supermartingale. For $S_n = 2^j$ and $j \geq 2$, we have similarly

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{n+1} v(S_{n+1})\right] &= \frac{1}{2}\left(\frac{4}{5}\right)^{n+1} v(2^{j+1}) + \frac{1}{2}\left(\frac{4}{5}\right)^{n+1} v(2^{j-1}) \\
&= \left(\frac{4}{5}\right)^n \left[\frac{2}{5} v(2^{j+1}) + \frac{2}{5} v(2^{j-1})\right] \\
&= \left(\frac{4}{5}\right)^n \left[\frac{2}{5} \cdot \frac{4}{2^{j+1}} + \frac{2}{5} \cdot \frac{4}{2^{j-1}}\right] \\
&= \left(\frac{4}{5}\right)^n \left[\frac{4}{5 \cdot 2^j} + \frac{16}{5 \cdot 2^j}\right] \\
&= \left(\frac{4}{5}\right)^n \cdot \frac{4}{2^j} = \left(\frac{4}{5}\right)^n v(S_n).
\end{aligned} \tag{5.4.7}$$

We thus see that in the "no exercise" region of stock prices 4 and above, the discounted option value is a martingale. It remains to consider the stock price $S_n = 2$ at the boundary between the "exercise" and "no exercise" regions. We compute

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\left(\frac{4}{5}\right)^{n+1} v(S_{n+1})\right] &= \frac{1}{2}\left(\frac{4}{5}\right)^{n+1} v(4) + \frac{1}{2}\left(\frac{4}{5}\right)^{n+1} v(1) \\
&= \left(\frac{4}{5}\right)^n \left[\frac{2}{5} v(4) + \frac{2}{5} v(1)\right] \\
&= \left(\frac{4}{5}\right)^n \left[\frac{2}{5} \cdot 1 + \frac{2}{5} \cdot 3\right] \\
&< \left(\frac{4}{5}\right)^n \cdot 2 = \left(\frac{4}{5}\right)^n v(S_n).
\end{aligned}$$

Again, we have a strict supermartingale. This means that if the stock price is 2, the value derived from immediate exercise is strictly greater than the value derived from the right to future exercise. The option should be exercised when the stock price is 2. The same is true when the stock price is less than 2.

Property (iii): Let Y_n, $n = 1, 2, \ldots$ be a process that satisfies

(a) $Y_n \geq (4 - S_n)^+$, $n = 0, 1, \ldots$;
(b) the discounted process $\left(\frac{4}{5}\right)^n Y_n$ is a supermartingale under the risk-neutral probability measure.

We must show that $v(S_n) \leq Y_n$ for all n. To do this, we fix n and consider two cases. If $S_n \leq 2$, then (5.4.6) and (a) imply that $v(S_n) = 4 - S_n \leq Y_n$. In the other case, $S_n = 2^j$ for some $j \geq 2$. In this case, we let τ denote the first time after time n that the stock price falls to the level 2. Using (5.4.5) but starting at time n at position S_n rather than at time 0, we have

$$v(S_n) = \widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{\tau - n} (4 - S_\tau)\right] = \widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{\tau - n} (4 - S_\tau)^+\right], \tag{5.4.8}$$

where the second equality follows from the fact that $4 - S_\tau = 2 > 0$. On the other hand, from (b), the Optional Sampling Theorem 4.3.2 of Chapter 4, and (a), we know that, for all $k \geq n$,

$$\begin{aligned}\left(\frac{4}{5}\right)^n Y_n &= \left(\frac{4}{5}\right)^{\tau\wedge n} Y_{\tau\wedge n}\\ &\geq \widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{\tau\wedge k} Y_{\tau\wedge k}\right]\\ &\geq \widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{\tau\wedge k}(4 - S_{\tau\wedge k})^+\right].\end{aligned} \tag{5.4.9}$$

Letting $k \to \infty$ in (5.4.9), we see that

$$\left(\frac{4}{5}\right)^n Y_n \geq \widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{\tau}(4 - S_\tau)^+\right].$$

Dividing by $\left(\frac{4}{5}\right)^n$, we obtain

$$Y_n \geq \widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{\tau-n}(4 - S_\tau)^+\right]. \tag{5.4.10}$$

Relations (5.4.8) and (5.4.10) imply that $v(S_n) \leq Y_n$.

Property (iii) guarantees that the buyer of the American perpetual put is not being overcharged if she pays $v(S_n)$ at time n when the stock price is S_n. The key equality in this proof when $S_n \geq 4$ is (5.4.8), which says that if the buyer uses the policy of exercising the first time the stock price falls to the level 2, the risk-neutral expected discounted payoff is exactly what she is paying for the put. Of course, if $S_n \leq 2$ and the buyer exercises immediately, she recovers the purchase price $v(S_n)$ of the put, which is the intrinsic value for these stock prices. Whenever we have a conjectured price for an American derivative security that is derived by evaluating some exercise policy, the buyer is not being overcharged (because she can always use this policy) and property (iii) will be satisfied.

The proof we have just given of the optimality of the policy of exercising the first time the stock price is at or below 2 is probabilistic in nature. It considers the stochastic process $v(S_n)$ and shows that, when discounted, this process is a supermartingale under the risk-neutral probabilities. There is a second method of proof, which is the discrete-time version of the partial differential equation characterization of the value function v for the perpetual American put. We give that now.

First note that we may rewrite (5.4.6) as

$$v(s) = \begin{cases} 4 - s & \text{if } s \leq 2,\\ \dfrac{4}{s} & \text{if } s \geq 4,\end{cases} \tag{5.4.11}$$

where, as before, we are really only interested in values of s of the form $s = 2^j$ for some integer j. We may recast conditions (i), (ii), and (iii) as

(i)′ $v(s) \geq (4 - s)^+$;

(ii)′ $v(s) \geq \frac{4}{5}\left[\frac{1}{2}v(2s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right]$;

(iii)′ $v(s)$ is the smallest function satisfying (i)′ and (ii)′. In other words, if $w(s)$ is another function satisfying (i)′ and (ii)′, then $v(s) \leq w(s)$ for every s of the form $s = 2^j$.

It is clear that (i)′ and (iii)′ are just restatements of (i) and (iii). Property (ii)′ implies

$$\widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{n+1} v(S_{n+1})\right] = \left(\frac{4}{5}\right)^n \frac{4}{5}\left[\frac{1}{2}v(2S_n) + \frac{1}{2}v\left(\frac{1}{2}S_n\right)\right] \leq \left(\frac{4}{5}\right)^n v(S_n),$$

which gives us property (ii).

If we have strict inequality in both (i)′ and (ii)′, then it is reasonable to expect that the function $v(s)$ could be made smaller and still satisfy (i)′ and (ii)′. This is indeed the case, although the proof is lengthy and will not be given. Because of property (iii)′, we cannot have strict inequality in both (i)′ and (ii)′. This gives us a fourth property:

(iv)′ For every s of the form $s = 2^j$, there is equality in either (i)′ or (ii)′.

In the case of $v(s)$ defined by (5.4.11), we have equality in (i)′ for $s \leq 2$. For $s \geq 4$, one can check, essentially as we did in (5.4.7), that equality holds in (ii)′:

$$\begin{aligned}\frac{4}{5}\left[\frac{1}{2}v(2s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right] &= \frac{2}{5}\cdot\frac{4}{2s} + \frac{2}{5}\cdot\frac{8}{s}\\ &= \frac{4}{5s} + \frac{16}{5s} = \frac{20}{5s} = \frac{4}{s} = v(s).\end{aligned}$$

Properties (i)′, (ii)′, and (iv)′ are conveniently summarized by the single equation

$$v(s) = \max\left\{(4 - s)^+, \frac{4}{5}\left[\frac{1}{2}v(2s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right]\right\}. \tag{5.4.12}$$

But for values of $s > 4$, which make $4 - s$ negative, $v(s)$ is equal to $\frac{4}{5}\left[\frac{1}{2}v(s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right]$, not $(4 - s)^+$. Therefore, for these values of s, it does not matter whether we write $(4 - s)^+$ or $4 - s$ in (5.4.12). For $s \leq 4$, the quantities $(4 - s)^+$ and $4 - s$ are the same. We may thus rewrite (5.4.12) more simply as

$$v(s) = \max\left\{4 - s, \frac{4}{5}\left[\frac{1}{2}v(2s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right]\right\}. \tag{5.4.13}$$

This is the so-called *Bellman equation* for the perpetual American put-pricing problem of this section.

The value $v(s)$ for the perpetual American put must satisfy (5.4.13), and this equation can be used to help determine the value. Unfortunately, there can be functions other than the one given by (5.4.11) that satisfy this equation. In particular, the function

$$w(s) = \frac{4}{s} \text{ for all } s = 2^j \tag{5.4.14}$$

also satisfies (5.4.13). The function $v(s)$ we seek is the *smallest* solution to (5.4.13) (see (iii)$'$). When we use (5.4.13) to solve for $v(s)$, we can use *boundary conditions* to rule out some of the extraneous solutions. In particular, the value of the perpetual American put must satisfy

$$\lim_{s \downarrow 0} v(s) = 4, \quad \lim_{s \to \infty} v(s) = 0. \tag{5.4.15}$$

The first of these conditions rules out the function $w(s)$ of (5.4.14).

In general, for a perpetual derivative security with intrinsic value $g(s)$, the value $v(s)$ of the option when the underlying stock price is s satisfies the Bellman equation

$$v(s) = \max\left\{g(s), \frac{1}{1+r}\left[\tilde{p}v(us) + \tilde{q}v(ds)\right]\right\}. \tag{5.4.16}$$

This is the same as equation (4.2.6) of Chapter 4 for the price of an American derivative security, except that when the derivative security is perpetual, its price does not depend on time; i.e., we have simply $v(s)$ rather than $v_n(s)$ and $v_{n+1}(s)$. If $g(s) = K - s$, so we are pricing a put with strike price K, then $v(s)$ should also satisfy the boundary conditions

$$\lim_{s \downarrow 0} v(s) = K, \quad \lim_{s \to \infty} v(s) = 0. \tag{5.4.17}$$

If $g(s) = K - s$, so we are pricing a call with strike price K, then $v(s)$ should satisfy the boundary conditions

$$\lim_{s \downarrow 0} v(s) = 0, \quad \lim_{s \to \infty} \frac{v(s)}{s} = 1. \tag{5.4.18}$$

(See Exercise 5.8.)

5.5 Summary

Let τ_m be the first time a symmetric random walk reaches level m, where m is a nonzero integer. Then

$$\mathbb{P}\{\tau_m < \infty\} = 1, \tag{5.2.12}$$

but

$$\mathbb{E}\tau_m = \infty. \tag{5.2.17}$$

If the random walk is asymmetric, with the probability p of an up step larger than $\frac{1}{2}$, then for m a positive integer, we have $\mathbb{P}\{\tau_m < \infty\} = 1$ and also $\mathbb{E}\tau_m < \infty$ (see Exercise 5.2 for the case of τ_1). If $0 < p < \frac{1}{2}$ and m is a positive integer, then $\mathbb{P}\{\tau_m = \infty\} > 0$ and $\mathbb{E}\tau_m = \infty$ (see Exercise 5.3 for the case of τ_1).

For both symmetric and asymmetric random walks, the moment generating function of τ_m can be computed (see (5.2.13) for the symmetric random walk, and see Exercises 5.2 and 5.3 for the asymmetric case). Knowledge of the moment-generating function permits us to determine the distribution of τ_m, although the computations are messy for large values of m. The distribution of τ_1 for the random walk with up-step probability p and down-step probability $q = 1 - p$ is

$$\mathbb{P}\{\tau_1 = 2j - 1\} = \frac{(2j-2)!}{j!(j-1)!}p^j q^{2j-1}, \quad j = 1, 2, \ldots. \tag{5.2.23}$$

See Exercise 5.4 for the case $m = 2$.

An alternative way to compute the distribution of τ_m is to use the *reflection principle*. The idea behind this principle is that there are as many paths that reach the level m prior to a final time but are below m at the final time as there are paths that are above m at the final time. This permits one to count the paths that reach level m by the final time. The reflection principle can also be used to determine the joint distribution of a random walk at some time n and the maximum value it achieves at or before time n; see Exercise 5.5.

The moment-generating function formula for a random walk permits us to evaluate the risk-neutral expected discounted payoff of a perpetual American put if it is exercised the first time the stock price falls to a specified threshold. One can then determine the exercise threshold that maximizes the risk-neutral expected discounted payoff. An example of this is worked out in Section 5.4. The argument that the put price found this way is indeed correct relies on verifying the three properties of the price of an American derivative security: (i) the price dominates the intrinsic value, (ii) the discounted price is a supermartingale under the risk-neutral measure, and (iii) the price obtained is the smallest possible price that satisfies (i) and (ii). These steps are carried out for the example in Section 5.4.

One can also seek the price of a perpetual American option by solving the Bellman equation, which for the put in Section 5.4 is

$$v(s) = \max\left\{4 - s, \frac{4}{5}\left[\frac{1}{2}v(2s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right]\right\}. \tag{5.4.13}$$

This equation must hold for all values of s that are on the lattice of possible stock prices. The continuous-time version of this equation studied in Chapter 8 of Volume II must hold for all positive values of s. In both discrete and continuous time, the Bellman equation can have extraneous solutions. In continuous time, one can use boundary conditions appropriate for the option under consideration to rule these out. In discrete time, one must use both the boundary conditions and a careful consideration of the nature of the lattice of possible stock prices to rule out the extraneous solutions. In particular, the continuous-time version of this equation for the general perpetual American put problem has a simpler solution than the discrete-time version.

5.6 Notes

The solution to the perpetual American put problem considered in Section 5.4 was worked out as a 1994 Carnegie Mellon Summer Undergraduate Mathematics Institute project by Irene Villegas using the method outlined in Exercise 5.9. This is an example of an optimal stopping problem. A classical reference for optimal stopping problems is Shiryaev [39]. Shiryaev, Kabanov, Kramkov, and Melnikov [41] consider the general perpetual American put problem in a binomial model, and their work is reported in Shiryaev [40].

5.7 Exercises

Exercise 5.1. For the symmetric random walk, consider the first passage time τ_m to the level m. The random variable $\tau_2 - \tau_1$ is the number of steps required for the random walk to rise from level 1 to level 2, and this random variable has the same distribution as τ_1, the number of steps required for the random walk to rise from level 0 to level 1. Furthermore, $\tau_2-\tau_1$ and τ_1 are independent of one another; the latter depends only on the coin tosses $1, 2, \dots, \tau_1$, and the former depends only on the coin tosses $\tau_1+1, \tau_1+2, \dots, \tau_2$.

(i) Use these facts to explain why

$$\mathbb{E}\alpha^{\tau_2} = (\mathbb{E}\alpha^{\tau_1})^2 \text{ for all } \alpha \in (0,1).$$

(ii) Without using (5.2.13), explain why for any positive integer m we must have

$$\mathbb{E}\alpha^{\tau_m} = (\mathbb{E}\alpha^{\tau_1})^m \text{ for all } \alpha \in (0,1). \tag{5.7.1}$$

(iii) Would equation (5.7.1) still hold if the random walk is not symmetric? Explain why or why not.

Exercise 5.2 (First passage time for random walk with upward drift). Consider the asymmetric random walk with probability p for an up step and probability $q = 1-p$ for a down step, where $\frac{1}{2} < p < 1$ so that $0 < q < \frac{1}{2}$. In the notation of (5.2.1), let τ_1 be the first time the random walk starting from level 0 reaches level 1. If the random walk never reaches this level, then $\tau_1 = \infty$.

(i) Define $f(\sigma) = pe^{\sigma} + qe^{-\sigma}$. Show that $f(\sigma) > 1$ for all $\sigma > 0$.

(ii) Show that, when $\sigma > 0$, the process

$$S_n = e^{\sigma M_n}\left(\frac{1}{f(\sigma)}\right)^n$$

is a martingale.

(iii) Show that, for $\sigma > 0$,

$$e^{-\sigma} = \mathbb{E}\left[\mathbb{I}_{\{\tau_1<\infty\}}\left(\frac{1}{f(\sigma)}\right)^{\tau_1}\right].$$

Conclude that $\mathbb{P}\{\tau_1 < \infty\} = 1$.

(iv) Compute $\mathbb{E}\alpha^{\tau_1}$ for $\alpha \in (0,1)$.

(v) Compute $\mathbb{E}\tau_1$.

Exercise 5.3 (First passage time for random walk with downward drift). Modify Exercise 5.2 by assuming $0 < p < \frac{1}{2}$ so that $\frac{1}{2} < q < 1$.

(i) Find a positive number σ_0 such that the function $f(\sigma) = pe^{\sigma} + qe^{-\sigma}$ satisfies $f(\sigma_0) = 1$ and $f(\sigma) > 1$ for all $\sigma > \sigma_0$.

(ii) Determine $\mathbb{P}\{\tau_1 < \infty\}$. (This quantity is no longer equal to 1.)

(iii) Compute $\mathbb{E}\alpha^{\tau_1}$ for $\alpha \in (0,1)$.

(iv) Compute $\mathbb{E}\left[\mathbb{I}_{\{\tau_1<\infty\}}\tau_1\right]$. (Since $\mathbb{P}\{\tau_1 = \infty\} > 0$, we have $\mathbb{E}\tau_1 = \infty$.)

Exercise 5.4 (Distribution of τ_2). Consider the symmetric random walk, and let τ_2 be the first time the random walk, starting from level 0, reaches the level 2. According to Theorem 5.2.3,

$$\mathbb{E}\alpha^{\tau_2} = \left(\frac{1-\sqrt{1-\alpha^2}}{\alpha}\right)^2 \text{ for all } \alpha \in (0,1).$$

Using the power series (5.2.21), we may write the right-hand side as

$$\begin{aligned}
\left(\frac{1-\sqrt{1-\alpha^2}}{\alpha}\right)^2 &= \frac{2}{\alpha}\cdot\frac{1-\sqrt{1-\alpha^2}}{\alpha} - 1\\
&= -1 + \sum_{j=1}^{\infty}\left(\frac{\alpha}{2}\right)^{2j-2}\frac{(2j-2)!}{j!(j-1)!}\\
&= \sum_{j=2}^{\infty}\left(\frac{\alpha}{2}\right)^{2j-2}\frac{(2j-2)!}{j!(j-1)!}\\
&= \sum_{k=1}^{\infty}\left(\frac{\alpha}{2}\right)^{2k}\frac{(2k)!}{(k+1)!k!}.
\end{aligned}$$

(i) Use the power series above to determine $\mathbb{P}\{\tau_2 = 2k\}$, $k = 1, 2, \ldots$.
(ii) Use the reflection principle to determine $\mathbb{P}\{\tau_2 = 2k\}$, $k = 1, 2, \ldots$.

Exercise 5.5 (Joint distribution of random walk and maximum-to-date). Let M_n be a symmetric random walk, and define its *maximum-to-date* process

$$M_n^* = \max_{1 \le k \le n} M_k. \tag{5.7.2}$$

Let n and m be even positive integers, and let b be an even integer less than or equal to m. Assume $m \le n$ and $2m - b \le n$.

(i) Use an argument based on reflected paths to show that

$$\begin{aligned}\mathbb{P}\{M_n^* \ge m, M_n = b\} &= \mathbb{P}\{M_n = 2m - b\} \\ &= \frac{n!}{\left(\frac{n-b}{2} + m\right)! \left(\frac{n+b}{2} - m\right)!} \left(\frac{1}{2}\right)^n.\end{aligned}$$

(ii) If the random walk is asymmetric with probability p for an up step and probability $q = 1 - p$ for a down step, where $0 < p < 1$, what is $\mathbb{P}\{M_n^* \ge m, M_n = b\}$?

Exercise 5.6. The value of the perpetual American put in Section 5.4 is the limit as $n \to \infty$ of the value of an American put with the same strike price 4 that expires at time n. When the initial stock price is $S_0 = 4$, the value of the perpetual American put is 1 (see (5.4.6) with $j = 2$). Show that the value of an American put in the same model when the initial stock price is $S_0 = 4$ is 0.80 if the put expires at time 1, 0.928 if the put expires at time 3, and 0.96896 if the put expires at time 5.

Exercise 5.7 (Hedging a short position in the perpetual American put). Suppose you have sold the perpetual American put of Section 5.4 and are hedging the short position in this put. Suppose that at the current time the stock price is s and the value of your hedging portfolio is $v(s)$. Your hedge is to first consume the amount

$$c(s) = v(s) - \frac{4}{5}\left[\frac{1}{2}v(2s) + v\left(\frac{s}{2}\right)\right] \tag{5.7.3}$$

and then take a position

$$\delta(s) = \frac{v(2s) - v\left(\frac{s}{2}\right)}{2s - \frac{s}{2}} \tag{5.7.4}$$

in the stock. (See Theorem 4.2.2 of Chapter 4. The processes C_n and Δ_n in that theorem are obtained by replacing the dummy variable s by the stock price S_n in (5.7.3) and (5.7.4); i.e., $C_n = c(S_n)$ and $\Delta_n = \delta(S_n)$.) If you hedge this way, then regardless of whether the stock goes up or down on the next step, the value of your hedging portfolio should agree with the value of the perpetual American put.

(i) Compute $c(s)$ when $s = 2^j$ for the three cases $j \le 0$, $j = 1$, and $j \ge 2$.
(ii) Compute $\delta(s)$ when $s = 2^j$ for the three cases $j \le 0$, $j = 1$, and $j \ge 2$.
(iii) Verify in each of the three cases $s = 2^j$ for $j \le 0$, $j = 1$, and $j \ge 2$ that the hedge works (i.e., regardless of whether the stock goes up or down, the value of your hedging portfolio at the next time is equal to the value of the perpetual American put at that time).

Exercise 5.8 (Perpetual American call). Like the perpetual American put of Section 5.4, the perpetual American call has no expiration. Consider a binomial model with up factor u, down factor d, and interest rate r that satisfies the no-arbitrage condition $0 < d < 1 + r < u$. The risk-neutral probabilities are

$$\tilde{p} = \frac{1 + r - d}{u - d}, \quad \tilde{q} = \frac{u - 1 - r}{u - d}.$$

The intrinsic value of the perpetual American call is $g(s) = s - K$, where $K > 0$ is the strike price. The purpose of this exercise is to show that the value of the call is always the price of the underlying stock, and there is no optimal exercise time.

(i) Let $v(s) = s$. Show that $v(S_n)$ is always at least as large as the intrinsic value $g(S_n)$ of the call and $\left(\frac{1}{1+r}\right)^n v(S_n)$ is a supermartingale under the risk-neutral probabilities. In fact, $\left(\frac{1}{1+r}\right)^n v(S_n)$ is a supermartingale. These are the analogues of properties (i) and (ii) for the perpetual American put of Section 5.4.
(ii) To show that $v(s) = s$ is not too large to be the value of the perpetual American call, we must find a good policy for the purchaser of the call. Show that if the purchaser of the call exercises at time n, regardless of the stock price at that time, then the discounted risk-neutral expectation of her payoff is $S_0 - \frac{K}{(1+r)^n}$. Because this is true for every n, and

$$\lim_{n \to \infty}\left[S_0 - \frac{K}{(1+r)^n}\right] = S_0,$$

the value of the call at time zero must be at least S_0. (The same is true at all other times; the value of the call is at least as great as the current stock price.)
(iii) In place of (i) and (ii) above, we could verify that $v(s) = s$ is the value of the perpetual American call by checking that this function satisfies the equation (5.4.16) and boundary conditions (5.4.18). Do this verification.
(iv) Show that there is no optimal time to exercise the perpetual American call.

Exercise 5.9. (Provided by Irene Villegas.) Here is a method for solving equation (5.4.13) for the value of the perpetual American put in Section 5.4.

(i) We first determine $v(s)$ for large values of s. When s is large, it is not optimal to exercise the put, so the maximum in (5.4.13) will be given by the second term,

$$\frac{4}{5}\left[\frac{1}{2}v(2s) + \frac{1}{2}v\left(\frac{s}{2}\right)\right] = \frac{2}{5}v(2s) + \frac{2}{5}v\left(\frac{s}{2}\right).$$

We thus seek solutions to the equation

$$v(s) = \frac{2}{5}v(2s) + \frac{2}{5}v\left(\frac{s}{2}\right). \tag{5.7.5}$$

All such solutions are of the form s^p for some constant p or linear combinations of functions of this form. Substitute s^p into (5.7.5), obtain a quadratic equation for 2^p, and solve to obtain $2^p = 2$ or $2^p = \frac{1}{2}$. This leads to the values $p = 1$ and $p = -1$, i.e., $v_1(s) = s$ and $v_2(s) = \frac{1}{s}$ are solutions to (5.7.5).

(ii) The general solution to (5.7.5) is a linear combination of $v_1(s)$ and $v_2(s)$, i.e.,

$$v(s) = As + \frac{B}{s}. \tag{5.7.6}$$

For large values of s, the value of the perpetual American put must be given by (5.7.6). It remains to evaluate A and B. Using the second boundary condition in (5.4.15), show that A must be zero.

(iii) We have thus established that for large values of s, $v(s) = \frac{B}{s}$ for some constant B still to be determined. For small values of s, the value of the put is its intrinsic value $4 - s$. We must choose B so these two functions coincide at some point, i.e., we must find a value for B so that, for some $s > 0$,

$$f_B(s) = \frac{B}{s} - (4 - s)$$

equals zero. Show that, when $B > 4$, this function does not take the value 0 for any $s > 0$, but, when $B \leq 4$, the equation $f_B(s) = 0$ has a solution.

(iv) Let B be less than or equal to 4, and let s_B be a solution of the equation $f_B(s) = 0$. Suppose s_B is a stock price that can be attained in the model (i.e., $s_B = 2^j$ for some integer j). Suppose further that the owner of the perpetual American put exercises the first time the stock price is s_B or smaller. Then the discounted risk-neutral expected payoff of the put is $v_B(S_0)$, where $v_B(s)$ is given by the formula

$$v_B(s) = \begin{cases} 4 - s, & \text{if } s \leq s_B, \\ \frac{B}{s}, & \text{if } s \geq s_B. \end{cases} \tag{5.7.7}$$

Which values of B and s_B give the owner the largest option value?

(v) For $s < s_B$, the derivative of $v_B(s)$ is $v_B'(s) = -1$. For $s > s_B$, this derivative is $v_B'(s) = -\frac{B}{s^2}$. Show that the best value of B for the option owner makes the derivative of $v_B(s)$ continuous at $s = s_B$ (i.e., the two formulas for $v_B'(s)$ give the same answer at $s = s_B$).

6

Interest-Rate-Dependent Assets

6.1 Introduction

In this chapter, we develop a simple, binomial model for interest rates and then examine some common assets whose value depends on interest rates. Assets in this class are called *fixed income* assets.

The simplest fixed income asset is a *zero-coupon bond*, a bond that pays a specified amount (called its *face value* or *par value*) at a specified time (called *maturity*). At times prior to maturity, the value of this asset is less than its face value, provided the interest rate is always greater than zero. One defines the *yield* of the zero-coupon bond corresponding to a given maturity as the constant interest rate that would be needed so that the time-zero price of the bond invested at time zero and allowed to accumulate at this interest rate would grow to the face value of the bond at the bond's maturity. Because there is theoretically a zero-coupon bond for each possible maturity, there is theoretically a yield corresponding to every time greater than the current time, which we call 0. The *yield curve* is the function from the maturity variable to the yield variable. One is interested in building models that not only determine the yield curve at a particular time but provide a method for random evolution of the yield curve forward in time. These are called *term structure of interest rates* models.

The existence of a zero-coupon bond for each maturity raises the specter of arbitrage. A term-structure model has numerous tradable assets—all the zero-coupon bonds and perhaps other fixed income assets. How can one be assured that one cannot within the model find an arbitrage by trading in these instruments? Regardless of one's view about the existence of arbitrage in the real world, it is clear that models built for pricing and hedging must not admit arbitrage. Any question that one attempts to answer with a model permitting arbitrage has a nonsensical answer. What is the price of an option on an asset? If one can begin with zero initial capital and hedge a short position in the option by using the arbitrage possibility within the model, one could argue that the option price is zero. Furthermore, one cannot advance a more convincing argument for any other price.

In this chapter, we avoid the problem of arbitrage by building the model under the risk-neutral measure. In other words, we first describe the evolution of the interest rate under the risk-neutral measure and then determine the prices of zero-coupon bonds and all other fixed income assets by using the risk-neutral pricing formula. This construction guarantees that all discounted asset prices are martingales, and so all discounted portfolio processes are martingales (when all coupons and other payouts of cash are reinvested in the portfolio). This is Theorem 6.2.6, the main result of Section 6.2. Because a martingale that begins at zero must always have expectation zero, martingale discounted portfolio processes are inconsistent with arbitrage; they cannot have a positive value with positive probability at some time unless they also have a negative value with positive probability at that time.

Models that are built starting from the interest rate and using risk-neutral pricing are called *short-rate models*. Examples of such models in continuous time are the Vasicek-Hull-White model [42], [23], and the Cox-Ingersoll-Ross model [10]. The classical model in continuous time that begins with a description of the random evolution of the yield curve and develops conditions under which the model is arbitrage-free is due to Heath, Jarrow, and Morton [20], [21]. This model might be called a *whole yield model* because it takes the whole yield curve as a starting point. The difference between these two classes of models is of fundamental practical importance. However, in a whole yield model, one still has an interest rate, and all the other assets are related to the interest rate by the risk-neutral pricing formula. Although the interest rate evolution in whole yield models is generally more complex than in the binomial model of this chapter, the ideas we develop here can be transferred to whole yield models.

Two common discrete-time models are Ho and Lee [22] and Black-Derman-Toy [4]. We use the former as the basis for Example 6.4.4 and the latter as the basis for Example 6.5.5.

6.2 Binomial Model for Interest Rates

Let Ω be the set of 2^N possible outcomes $\omega_1\omega_2\dots\omega_N$ of N tosses of a coin, and let $\widetilde{\mathbb{P}}$ be a probability measure on Ω under which every sequence $\omega_1\omega_2\dots\omega_N$ has strictly positive probability. We define an *interest rate process* to be a sequence of random variables

$$R_0, R_1, \dots, R_{N-1},$$

where R_0 is not random and, for $n = 1, \dots, N-1$, R_n depends only on the first n coin tosses $\omega_1 \dots \omega_n$. One dollar invested in the money market at time n grows to $1 + R_n$ dollars at time $n+1$. Although the interest rate is random, at time n we know the interest rate that will be applied to money market investments over the period from time n to time $n+1$. This is less random than a stock. If we invest in a stock at time n, we do not know what the value of the investment will be at time $n+1$.

It is natural to assume that $R_n > 0$ for all n and all $\omega_1 \dots \omega_n$. This is the case to keep in mind. However, the analysis requires only that

$$R_n(\omega_1 \dots \omega_n) > -1 \text{ for all } n \text{ and all } \omega_1, \dots, \omega_n, \tag{6.2.1}$$

and this is the only assumption we make.

We define the *discount process* by

$$D_n = \frac{1}{(1+R_0)\cdots(1+R_{n-1})}, \quad n = 1, 2, \dots, N; \quad D_0 = 1. \tag{6.2.2}$$

Note that D_n depends on only the first $n-1$ coin tosses, contrary to the usual situation in which the subscript indicates the number of coin tosses on which the random variable depends.

The risk-neutral pricing formula says that the value at time zero of a payment X received at time m (where X is allowed to depend only on $\omega_1 \dots \omega_m$) is

$$\widetilde{\mathbb{E}}[D_m X],$$

the risk-neutral expected discounted payment. We use this formula to define the time-zero price of a zero-coupon bond that pays 1 at maturity time m to be

$$B_{0,m} = \widetilde{\mathbb{E}}[D_m]. \tag{6.2.3}$$

The yield for this bond is the number y_m for which

$$\frac{1}{B_{0,m}} = (1 + y_m)^m.$$

At time zero, an investment of one dollar in the m-maturity bond would purchase $\frac{1}{B_{0,m}}$ of these bonds, and this investment would pay off $\frac{1}{B_{0,m}}$ at time m. This is the same as investing one dollar at a constant rate of interest y_m between times 0 and m. Of course, we can solve the equation above for y_m:

$$y_m = \left(\frac{1}{B_{0,m}}\right)^{\frac{1}{m}} - 1.$$

We would also like to define the price of the m-maturity zero-coupon bond at time n, where $1 \le n \le m$. To do that, we need to extend the notion of conditional expectation of Definition 2.3.1. Consider the following example.

Example 6.2.1. Assume $N = 3$ so that

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Assume further that

$$\widetilde{\mathbb{P}}\{HHH\} = \frac{2}{9}, \quad \widetilde{\mathbb{P}}\{HHT\} = \frac{1}{9}, \quad \widetilde{\mathbb{P}}\{HTH\} = \frac{1}{12}, \quad \widetilde{\mathbb{P}}\{HTT\} = \frac{1}{12}$$
$$\widetilde{\mathbb{P}}\{THH\} = \frac{1}{6}, \quad \widetilde{\mathbb{P}}\{THT\} = \frac{1}{12}, \quad \widetilde{\mathbb{P}}\{TTH\} = \frac{1}{8}, \quad \widetilde{\mathbb{P}}\{TTT\} = \frac{1}{8}.$$

These numbers sum to 1. We define the sets

$$\begin{aligned} A_{HH} &= \{\omega_1 = H, \omega_2 = H\} = \{HHH, HHT\}, \\ A_{HT} &= \{\omega_1 = H, \omega_2 = T\} = \{HTH, HTT\}, \\ A_{TH} &= \{\omega_1 = T, \omega_2 = H\} = \{THH, THT\}, \\ A_{TT} &= \{\omega_1 = T, \omega_2 = T\} = \{TTH, TTT\} \end{aligned}$$

of outcomes that begin with HH, HT, TH, and TT on the first two tosses, respectively. We have

$$\widetilde{\mathbb{P}}\{A_{HH}\} = \frac{1}{3}, \quad \widetilde{\mathbb{P}}\{A_{HT}\} = \frac{1}{6}, \quad \widetilde{\mathbb{P}}\{A_{TH}\} = \frac{1}{4}, \quad \widetilde{\mathbb{P}}\{A_{TT}\} = \frac{1}{4}.$$

Similarly, we define the sets

$$\begin{aligned} A_H &= \{\omega_1 = H\} = \{HHH, HHT, HTH, HTT\}, \\ A_T &= \{\omega_1 = T\} = \{THH, THT, TTH, TTT\} \end{aligned}$$

of outcomes that begin with H and T on the first toss, respectively. For these sets, we have

$$\widetilde{\mathbb{P}}\{A_H\} = \frac{1}{2}, \quad \widetilde{\mathbb{P}}\{A_T\} = \frac{1}{2}.$$

If the coin is tossed three times and we are told only that the first two tosses are HH, then we know that the outcome of the three tosses is in A_{HH}. Conditioned on this information, the probability that the third toss is an H is

$$\widetilde{\mathbb{P}}\{\omega_3 = H | \omega_1 = H, \omega_2 = H\} = \frac{\widetilde{\mathbb{P}}\{HHH\}}{\widetilde{\mathbb{P}}\{A_{HH}\}} = \frac{2/9}{1/3} = \frac{2}{3}.$$

Similarly, we have

$$\widetilde{\mathbb{P}}\{\omega_3 = T | \omega_1 = H, \omega_2 = H\} = \frac{\widetilde{\mathbb{P}}\{HHT\}}{\widetilde{\mathbb{P}}\{A_{HH}\}} = \frac{1/9}{1/3} = \frac{1}{3},$$
$$\widetilde{\mathbb{P}}\{\omega_3 = H | \omega_1 = H, \omega_2 = T\} = \frac{\widetilde{\mathbb{P}}\{HTH\}}{\widetilde{\mathbb{P}}\{A_{HT}\}} = \frac{1/12}{1/6} = \frac{1}{2},$$
$$\widetilde{\mathbb{P}}\{\omega_3 = T | \omega_1 = H, \omega_2 = T\} = \frac{\widetilde{\mathbb{P}}\{HTT\}}{\widetilde{\mathbb{P}}\{A_{HT}\}} = \frac{1/12}{1/6} = \frac{1}{2},$$

$$\widetilde{\mathbb{P}}\{\omega_3 = H | \omega_1 = T, \omega_2 = H\} = \frac{\widetilde{\mathbb{P}}\{THH\}}{\widetilde{\mathbb{P}}\{A_{TH}\}} = \frac{1/6}{1/4} = \frac{2}{3},$$
$$\widetilde{\mathbb{P}}\{\omega_3 = T | \omega_1 = T, \omega_2 = H\} = \frac{\widetilde{\mathbb{P}}\{THT\}}{\widetilde{\mathbb{P}}\{A_{TH}\}} = \frac{1/12}{1/4} = \frac{1}{3},$$
$$\widetilde{\mathbb{P}}\{\omega_3 = H | \omega_1 = T, \omega_2 = T\} = \frac{\widetilde{\mathbb{P}}\{TTH\}}{\widetilde{\mathbb{P}}\{A_{TT}\}} = \frac{1/8}{1/4} = \frac{1}{2},$$
$$\widetilde{\mathbb{P}}\{\omega_3 = T | \omega_1 = T, \omega_2 = T\} = \frac{\widetilde{\mathbb{P}}\{TTT\}}{\widetilde{\mathbb{P}}\{A_{TT}\}} = \frac{1/8}{1/4} = \frac{1}{2}.$$

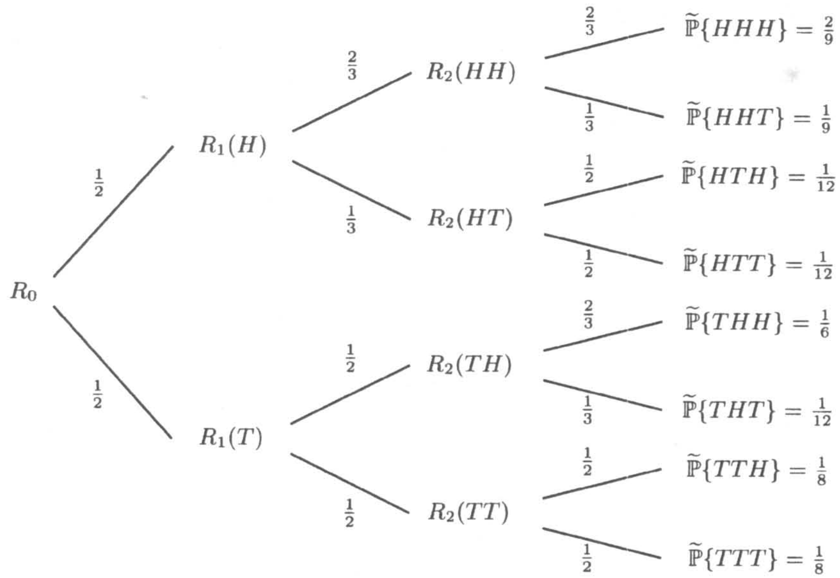

Fig. 6.2.1. A three-period interest rate model.

If the coin is tossed three times and we are told only that the first toss is H, then we know that the outcome of the three tosses is in A_H. Conditioned on this information, the probability that the second toss is H is

$$\widetilde{\mathbb{P}}\{\omega_2 = H | \omega_1 = H\} = \frac{\widetilde{\mathbb{P}}\{A_{HH}\}}{\widetilde{\mathbb{P}}\{A_H\}} = \frac{1/3}{1/2} = \frac{2}{3}.$$

Similarly,

$$\widetilde{\mathbb{P}}\{\omega_2 = T|\omega_1 = H\} = \frac{\widetilde{\mathbb{P}}\{A_{HT}\}}{\widetilde{\mathbb{P}}\{A_H\}} = \frac{1/6}{1/2} = \frac{1}{3},$$

$$\widetilde{\mathbb{P}}\{\omega_2 = H|\omega_1 = T\} = \frac{\widetilde{\mathbb{P}}\{A_{TH}\}}{\widetilde{\mathbb{P}}\{A_T\}} = \frac{1/4}{1/2} = \frac{1}{2},$$

$$\widetilde{\mathbb{P}}\{\omega_2 = T|\omega_1 = T\} = \frac{\widetilde{\mathbb{P}}\{A_{TT}\}}{\widetilde{\mathbb{P}}\{A_T\}} = \frac{1/4}{1/2} = \frac{1}{2}.$$

These conditional probabilities are the probabilities of traversing the indicated links in the tree shown in Figure 6.2.1. We shall call them the *transition probabilities* associated with $\widetilde{\mathbb{P}}$. The (unconditional) probabilities of following each of the eight paths are shown at the right ends of the paths.

We may compute $\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = H\}$ by two methods, either computing along the upper links in Figure 6.2.1 using the formula

$$\begin{aligned}
&\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = H\}\\
&= \widetilde{\mathbb{P}}\{\omega_2 = H|\omega_1 = H\} \cdot \widetilde{\mathbb{P}}\{\omega_3 = H|\omega_1 = H, \omega_2 = H\}\\
&= \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}
\end{aligned}$$

or by using the formula

$$\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = H\} = \frac{\widetilde{\mathbb{P}}\{HHH\}}{\widetilde{\mathbb{P}}(A_H)} = \frac{2/9}{1/2} = \frac{4}{9}.$$

By either of these two methods, we have the following formulas:

$$\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = H\} = \frac{4}{9}, \quad \widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = T|\omega_1 = H\} = \frac{2}{9},$$

$$\widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = H|\omega_1 = H\} = \frac{1}{6}, \quad \widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = T|\omega_1 = H\} = \frac{1}{6},$$

$$\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = T\} = \frac{1}{3}, \quad \widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = T|\omega_1 = T\} = \frac{1}{6},$$

$$\widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = H|\omega_1 = T\} = \frac{1}{4}, \quad \widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = T|\omega_1 = T\} = \frac{1}{4}.$$

Finally, suppose we have a random variable $X(\omega_1\omega_2\omega_3)$. We define the conditional expectation of this random variable, given the first coin toss, by the formulas

$$\begin{aligned}
\widetilde{\mathbb{E}}_1[X](H) = {} & X(HHH)\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = H\}\\
&+X(HHT)\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = T|\omega_1 = H\}\\
&+X(HTH)\widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = H|\omega_1 = H\}\\
&+X(HTT)\widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = T|\omega_1 = H\}\\
= {} & \frac{4}{9}X(HHH) + \frac{2}{9}X(HHT) + \frac{1}{6}X(HTH) + \frac{1}{6}X(HTT),
\end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbb{E}}_1[X](T) = {} & X(THH)\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = H|\omega_1 = T\}\\
&+X(THT)\widetilde{\mathbb{P}}\{\omega_2 = H, \omega_3 = T|\omega_1 = T\}\\
&+X(TTH)\widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = H|\omega_1 = T\}\\
&+X(TTT)\widetilde{\mathbb{P}}\{\omega_2 = T, \omega_3 = T|\omega_1 = T\}\\
= {} & \frac{1}{3}X(THH) + \frac{1}{6}X(THT) + \frac{1}{4}X(TTH) + \frac{1}{4}X(TTT).
\end{aligned}$$

Note that $\widetilde{\mathbb{E}}_1[X]$ is a random variable because it depends on the first toss. We define the conditional expectation of X, given the first two tosses, by the formulas

$$\begin{aligned}
\widetilde{\mathbb{E}}_2[X](HH) &= X(HHH)\widetilde{\mathbb{P}}\{\omega_3 = H|\omega_1 = H, \omega_2 = H\}\\
&\quad + X(HHT)\widetilde{\mathbb{P}}\{\omega_3 = T|\omega_1 = H, \omega_2 = H\}\\
&= \frac{2}{3}X(HHH) + \frac{1}{3}X(HHT),
\end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbb{E}}_2[X](HT) &= X(HTH)\widetilde{\mathbb{P}}\{\omega_3 = H|\omega_1 = H, \omega_2 = T\}\\
&\quad + X(HTT)\widetilde{\mathbb{P}}\{\omega_3 = T|\omega_1 = H, \omega_2 = T\}\\
&= \frac{1}{2}X(HTH) + \frac{1}{2}X(HTT),
\end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbb{E}}_2[X](TH) &= X(THH)\widetilde{\mathbb{P}}\{\omega_3 = H|\omega_1 = T, \omega_2 = H\}\\
&\quad + X(THT)\widetilde{\mathbb{P}}\{\omega_3 = T|\omega_1 = T, \omega_2 = H\}\\
&= \frac{2}{3}X(THH) + \frac{1}{3}X(THT),
\end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbb{E}}_2[X](TT) &= X(TTH)\widetilde{\mathbb{P}}\{\omega_3 = H|\omega_1 = T, \omega_2 = T\}\\
&\quad + X(TTT)\widetilde{\mathbb{P}}\{\omega_3 = T|\omega_1 = T, \omega_2 = T\}\\
&= \frac{1}{2}X(TTH) + \frac{1}{2}X(TTT).
\end{aligned}$$

Note that $\widetilde{\mathbb{E}}_2[X]$ is also a random variable; it depends on the first two tosses.

If the random variable X whose conditional expectation we are computing depends on only the first two tosses, then the computations above simplify. In particular, we would have

$$\begin{aligned}
\widetilde{\mathbb{E}}_1[X](H) &= \left(\frac{4}{9} + \frac{2}{9}\right)X(HH) + \left(\frac{1}{6} + \frac{1}{6}\right)X(HT)\\
&= \frac{2}{3}X(HH) + \frac{1}{3}X(HT),
\end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbb{E}}_1[X](T) &= \left(\frac{1}{3} + \frac{1}{6}\right)X(TH) + \left(\frac{1}{4} + \frac{1}{4}\right)X(TT)\\
&= \frac{1}{2}X(TH) + \frac{1}{2}X(TT),
\end{aligned}$$

and

$$\widetilde{\mathbb{E}}_2[X](HH) = X(HH), \ \widetilde{\mathbb{E}}_2[X](HT) = X(HT),$$
$$\widetilde{\mathbb{E}}_2[X](TH) = X(TH), \ \widetilde{\mathbb{E}}_2[X](TT) = X(TT). \qquad \square$$

The previous example is a special case of the following definition.

Definition 6.2.2. *Let $\widetilde{\mathbb{P}}$ be a probability measure on the space Ω of all possible sequences of N coin tosses. Assume that every sequence $\omega_1 \dots \omega_N$ in Ω has positive probability under $\widetilde{\mathbb{P}}$. Let $1 \le n \le N-1$, and let $\overline{\omega}_1 \dots \overline{\omega}_N$ be a sequence of N coin tosses. We define*

$$\widetilde{\mathbb{P}}\{\omega_{n+1} = \overline{\omega}_{n+1}, \dots, \omega_N = \overline{\omega}_N | \omega_1 = \overline{\omega}_1, \dots, \omega_n = \overline{\omega}_n\} = \frac{\widetilde{\mathbb{P}}\{\overline{\omega}_1 \dots \overline{\omega}_n \overline{\omega}_{n+1} \dots \overline{\omega}_N\}}{\widetilde{\mathbb{P}}\{\omega_1 = \overline{\omega}_1, \dots, \omega_n = \overline{\omega}_n\}}.$$

Let X be a random variable. We define the conditional expectation of X based on the information at time n *by the formula*

$$\widetilde{\mathbb{E}}_n[X](\overline{\omega}_1 \dots \overline{\omega}_n) = \sum_{\overline{\omega}_{n+1}, \dots, \overline{\omega}_N} X(\overline{\omega}_1 \dots \overline{\omega}_n \overline{\omega}_{n+1} \dots \overline{\omega}_N) \times \widetilde{\mathbb{P}}\{\omega_{n+1} = \overline{\omega}_{n+1}, \dots, \omega_N = \overline{\omega}_N | \omega_1 = \overline{\omega}_1, \dots, \omega_n = \overline{\omega}_n\}.$$

We further set $\widetilde{\mathbb{E}}_0[X] = \widetilde{\mathbb{E}}X$ and $\widetilde{\mathbb{E}}_N[X] = X$.

Remark 6.2.3. Conditional expectation as defined above is a generalization of Definition 2.3.1 that allows for nonindependent coin tosses. It satisfies all the properties of Theorem 2.3.2. The proofs of these properties are straightforward modifications of the proofs given in the appendix.

Definition 6.2.4 (Zero-coupon bond prices). *Let $\widetilde{\mathbb{P}}$ be a probability measure on the space Ω of all possible sequences of N coin tosses, and assume that every sequence $\omega_1 \dots \omega_N$ has strictly positive probability under $\widetilde{\mathbb{P}}$. Let $R_0, R_1, R_2, \dots, R_{N-1}$ be an interest rate process, with each R_n depending on only the first n coin tosses and satisfying (6.2.1). Define the discount process D_n, $n = 0, 1, \dots, N$, by (6.2.2). For $0 \le n \le m \le N$, the* price at time n of the zero-coupon bond maturing at time m *is defined to be*

$$B_{n,m} = \widetilde{\mathbb{E}}_n\left[\frac{D_m}{D_n}\right]. \tag{6.2.4}$$

According to (6.2.4), the price at time m of the zero-coupon bond maturing at time m is $B_{m,m} = 1$. We are choosing the face value 1 as a convenient normalization. If the face value of a bond is some other quantity C, then the value of this bond at time n before maturity m is $CB_{n,m}$. If $n = 0$, then (6.2.4) reduces to (6.2.3).

Remark 6.2.5. We may use the property of "taking out what is known" (Theorem 2.3.2(ii)) to rewrite (6.2.4) as

$$D_n B_{n,m} = \widetilde{\mathbb{E}}_n[D_m]. \tag{6.2.5}$$

From this formula, we see that the discounted bond price $D_n B_{n,m}$, $n = 0, 1, \dots, m$, is a martingale under $\widetilde{\mathbb{P}}$. In particular, if $0 \le k \le n \le m$, then (6.2.5) and "iterated conditioning" (Theorem 2.3.2(iii)) imply

$$\widetilde{\mathbb{E}}_k[D_n B_{n,m}] = \widetilde{\mathbb{E}}_k\big[\widetilde{\mathbb{E}}_n[D_m]\big] = \widetilde{\mathbb{E}}_k[D_m] = D_k B_{k,m},$$

which is the martingale property. Definition 6.2.4 was chosen so that discounted zero-coupon bond prices would be martingales.

Consider an agent who can trade in the zero-coupon bonds of every maturity and in the money market. We wish to show that the wealth of such an agent, when discounted, constitutes a martingale. From this, we shall see that there is no arbitrage in the model in which zero-coupon bond prices are defined by (6.2.4). In particular, let $\Delta_{n,m}$ be the number of m-maturity zero-coupon bonds held by the agent between times n and $n+1$. Of course, we must have $n < m$ since the m-maturity bond does not exist after time m. Furthermore, $\Delta_{n,m}$ is allowed to depend on only the first n coin tosses. The agent begins with a nonrandom initial wealth X_0 and at each time n has wealth X_n. His wealth at time $n+1$ is thus given by

$$X_{n+1} = \Delta_{n,n+1} + \sum_{m=n+2}^{N} \Delta_{n,m} B_{n+1,m} + (1+R_n)\left(X_n - \sum_{m=n+1}^{N} \Delta_{n,m} B_{n,m}\right). \tag{6.2.6}$$

The first term on the right-hand side of (6.2.6) is the payoff of the zero-coupon bond maturing at time $n+1$, multiplied by the positions taken in this bond at time n and held to time $n+1$. The second term is the value of all the zero-coupon bonds maturing at time $n+2$ and later, multiplied by the position taken in these bonds at time n and held to $n+1$. The second factor in the third term is the cash position taken at time n, which is the difference between the total wealth X_n of the agent and the value of all the bonds held after rebalancing at time n. This is multiplied by 1 plus the interest rate that prevails between times n and $n+1$.

Theorem 6.2.6. *Regardless of how the portfolio random variables $\Delta_{n,m}$ are chosen (subject to the condition that $\Delta_{n,m}$ may depend only on the first n coin tosses), the discounted wealth process $D_n X_n$ is a martingale under $\widetilde{\mathbb{P}}$.*

PROOF: We use the fact that $\Delta_{n,m}$ depends on only the first n coin tosses and "taking out what is known" (Theorem 2.3.2(ii)), the fact that D_{n+1} also depends only on the first n coin tosses, and the martingale property of Remark 6.2.5 to compute

$$\begin{aligned}
\widetilde{\mathbb{E}}_n[X_{n+1}] &= \Delta_{n,n+1} + \sum_{m=n+2}^{N} \Delta_{n,m}\widetilde{\mathbb{E}}_n[B_{n+1,m}] \\
&\quad +(1+R_n)\left(X_n - \sum_{m=n+1}^{N} \Delta_{n,m}B_{n,m}\right) \\
&= \Delta_{n,n+1} + \sum_{m=n+2}^{N} \frac{\Delta_{n,m}}{D_{n+1}}\widetilde{\mathbb{E}}_n[D_{n+1}B_{n+1,m}] \\
&\quad +\frac{D_n}{D_{n+1}}\left(X_n - \sum_{m=n+1}^{N} \Delta_{n,m}B_{n,m}\right) \\
&= \Delta_{n,n+1} + \sum_{m=n+2}^{N} \frac{\Delta_{n,m}}{D_{n+1}} D_n B_{n,m} + \frac{D_n}{D_{n+1}}X_n \\
&\quad -\frac{D_n}{D_{n+1}} \sum_{m=n+1}^{N} \Delta_{n,m}B_{n,m} \\
&= \Delta_{n,n+1} + \frac{D_n}{D_{n+1}}X_n - \frac{D_n}{D_{n+1}}\Delta_{n,n+1}B_{n,n+1}.
\end{aligned}$$

But $B_{n,n+1} = \widetilde{\mathbb{E}}_n\left[\frac{D_{n+1}}{D_n}\right] = \frac{D_{n+1}}{D_n}$ because $\frac{D_{n+1}}{D_n}$ depends on only the first n coin tosses. Substituting this into the equation above, we obtain

$$\widetilde{\mathbb{E}}_n[X_{n+1}] = \frac{D_n}{D_{n+1}}X_n.$$

Using again the fact that D_{n+1} depends only on the first n coin tosses, we may rewrite this as

$$\widetilde{\mathbb{E}}_n[D_{n+1}X_{n+1}] = D_nX_n,$$

which is the martingale property for the discounted wealth process. □

Remark 6.2.7. Because the discounted wealth process is a martingale under $\widetilde{\mathbb{P}}$, it has constant expectation:

$$\widetilde{\mathbb{E}}[D_nX_n] = X_0,\ n = 0, 1, \dots, N. \tag{6.2.7}$$

If one could construct an arbitrage by trading in the zero-coupon bonds and the money market, then there would be a portfolio process that begins with $X_0 = 0$ and at some future time n results in $X_n \geq 0$, regardless of the outcome of the coin tossing, and further results in $X_n > 0$ for some of the outcomes. In such a situation, we would have $\widetilde{\mathbb{E}}[D_nX_n] > 0 = X_0$, a situation that is ruled out by (6.2.7). In other words, by using the risk-neutral pricing formula (6.2.4) to define zero-coupon bond prices, we have succeeded in building a model that is free of arbitrage.

Remark 6.2.8. The risk-neutral pricing formula says that for $0 \leq n \leq m \leq N$, the price at time n of a derivative security paying V_m at time m (where V_m depends on only the first m coin tosses) is

$$V_n = \frac{1}{D_n}\widetilde{\mathbb{E}}_n[D_mV_m]. \tag{6.2.8}$$

Theorem 6.2.6 provides a partial justification for this. Namely, if it is possible to construct a portfolio that hedges a short position in the derivative security (i.e., that has value V_m at time m regardless of the outcome of the coin tossing), then the value of the derivative security at time n must be V_n given by (6.2.8). Theorem 6.2.6 does not guarantee that such a hedging portfolio can be constructed.

Remark 6.2.9. In the wealth equation (6.2.6), an agent is permitted to invest in the money market and zero-coupon bonds of all maturities. However, investing at time n in the zero-coupon bond with maturity $n+1$ is the same as investing in the money market. An investment of 1 in this bond at time n purchases $\frac{1}{B_{n,n+1}}$ bonds maturing at time $n+1$, and this investment pays off $\frac{1}{B_{n,n+1}}$ at time $n+1$. But

$$\frac{1}{B_{n,n+1}} = \frac{1}{\widetilde{\mathbb{E}}_n\left[\frac{D_{n+1}}{D_n}\right]} = \frac{1}{\widetilde{\mathbb{E}}_n\left[\frac{1}{1+R_n}\right]} = \frac{1}{\frac{1}{1+R_n}} = 1 + R_n.$$

This is the same payoff one would receive by investing 1 in the money market at time n. It is convenient to have a money market account and write (6.2.6) as we did, but it is actually unnecessary to include the money market account among the traded assets in this discrete-time model.

Let $0 \leq m \leq N$ be given. A *coupon-paying bond* can be modeled as a sequence of constant (i.e., nonrandom) quantities $C_0, C_1, \dots, C_m$. For $0 \leq n \leq m-1$, the constant C_n is the coupon payment made at time n (which may be zero). The constant C_m is the payment made at time m, which includes principal as well as any coupon due at time m. In the case of the zero-coupon bond of Definition 6.2.4, $C_0 = C_1 = \cdots = C_{m-1} = 0$ and $C_m = 1$. In general, we may regard a coupon-paying bond as a sum of C_1 zero-coupon bonds maturing at time 1, C_2 zero-coupon bonds maturing at time 2, etc., up to C_m zero-coupon bonds maturing at time m. The price at time zero of the coupon-paying bond is thus

$$\sum_{k=0}^{m} C_kB_{0,k} = \widetilde{\mathbb{E}}\left[\sum_{k=0}^{m} D_kC_k\right].$$

At time n, before the payment C_n has been made but after the payments $C_0, \dots, C_{n-1}$ have been made, the price of the coupon-paying bond is

$$\sum_{k=n}^{m} C_kB_{n,k} = \widetilde{\mathbb{E}}_n\left[\sum_{k=n}^{m} \frac{C_kD_k}{D_n}\right],\ n = 0, 1, \dots, m. \tag{6.2.9}$$

This generalizes formula (2.4.13) of Chapter 2 to the case of nonconstant interest rates but constant payment quantities $C_0, C_1, \ldots, C_m$.

6.3 Fixed-Income Derivatives

Suppose that in the binomial model for interest rates we have an asset whose price at time n we denote by S_n. This may be a stock but is more often a contract whose payoff depends on the interest rate. The price S_n is allowed to depend on the first n coin tosses. We have taken $\widetilde{\mathbb{P}}$ to be a risk-neutral measure, which means that, under $\widetilde{\mathbb{P}}$, the discounted asset price is a martingale:

$$D_n S_n = \widetilde{\mathbb{E}}_n\left[D_{n+1} S_{n+1}\right], \quad n = 0, 1, \ldots, N-1. \tag{6.3.1}$$

Definition 6.3.1. *A* forward contract *is an agreement to pay a specified* delivery price K *at a* delivery date m, *where* $0 \le m \le N$, *for the asset whose price at time* m *is* S_m. *The* m-forward price *of this asset at time* n, *where* $0 \le n \le m$, *is the value of* K *that makes the forward contract have no-arbitrage price zero at time* n.

Theorem 6.3.2. *Consider an asset with price process* $S_0, S_1, \ldots, S_N$ *in the binomial interest rate model. Assume that zero-coupon bonds of all maturities can be traded. For* $0 \le n \le m \le N$, *the* m*-forward price at time* n *of the asset is*

$$\mathrm{For}_{n,m} = \frac{S_n}{B_{n,m}}. \tag{6.3.2}$$

PROOF: Suppose at time n an agent sells the forward contract with delivery date m and delivery price K. Suppose further that the value of K is chosen so that the forward contract has price zero at time n. Then, selling the forward contract generates no income. Having sold the forward contract at time n, suppose an agent immediately shorts $\frac{S_n}{B_{n,m}}$ zero-coupon bonds and uses the income S_n generated to buy one share of the asset. The agent then does no further trading until time m, at which time he owns one share of the asset, which he delivers according to the forward contract. In exchange, he receives K. After covering the short bond position, he is left with $K - \frac{S_n}{B_{n,m}}$. If this is positive, the agent has found an arbitrage. If it is negative, the agent could instead have taken the opposite positions, going long the forward, long the m-maturity bond, and short the asset, to again achieve an arbitrage. In order to preclude arbitrage, K must be given by (6.3.2). □

Remark 6.3.3. The proof of Theorem 6.3.2 constructs the hedge for a short position in a forward contract. This is called a *static hedge* because it calls for no trading between the time n when the hedge is set up and the time m when the forward contract expires. Whenever there is a hedge, static or not, the discounted value of the hedging portfolio is a martingale under the

risk-neutral measure, and hence the risk-neutral pricing formula applies. In this case, that formula says the forward contract whose payoff is $S_m - K$ at time m must have time-n discounted price

$$\widetilde{\mathbb{E}}_n[D_m(S_m - K)] = \widetilde{\mathbb{E}}_n[D_m S_m] - K D_n \widetilde{\mathbb{E}}_n\left[\frac{D_m}{D_n}\right] = D_n\left(S_n - K B_{n,m}\right),$$

where we have used the martingale property (6.3.1) and the definition (6.2.4) of $B_{n,m}$. In order for the time-n price of the forward contract to be zero, we must have $S_n - K B_{n,m} = 0$, i.e., K is given by (6.3.2).

In addition to forward prices of assets, one can define forward interest rates. Let $0 \le n \le m \le N-1$ be given. Suppose at time n an agent shorts an m-maturity zero-coupon bond and uses the income generated by this to purchase $\frac{B_{n,m}}{B_{n,m+1}}$ zero-coupon bonds maturing at time $m+1$. This portfolio is set up at zero cost at time n. At time m, it requires the agent to invest 1 to cover the short position in the m-maturity bond. At time $m+1$, the agent receives $\frac{B_{n,m}}{B_{n,m+1}}$ from the long position in the $m+1$-maturity bonds. Thus, between times m and $m+1$, it is as if the agent has invested at the interest rate

$$F_{n,m} = \frac{B_{n,m}}{B_{n,m+1}} - 1 = \frac{B_{n,m} - B_{n,m+1}}{B_{n,m+1}}. \tag{6.3.3}$$

Moreover, this interest rate for investing between times m and $m+1$ was "locked in" by the portfolio set up at time n. Note that

$$F_{m,m} = \frac{B_{m,m}}{B_{m,m+1}} - 1 = \frac{1}{\frac{1}{1+R_m}} - 1 = R_m.$$

The interest rate that can be locked in at time m for borrowing at time m is R_m.

Definition 6.3.4. *Let* $0 \le n \le m \le N-1$ *be given. The* forward interest rate *at time* n *for investing at time* m *is defined by (6.3.3).*

Theorem 6.3.5. *Let* $0 \le n \le m \le N-1$ *be given. The no-arbitrage price at time* n *of a contract that pays* R_m *at time* $m+1$ *is* $B_{n,m+1}F_{n,m} = B_{n,m} - B_{n,m+1}$.

PROOF: Suppose at time n an agent sells a contract that promises to pay R_m at time $m+1$ and receives income $B_{n,m} - B_{n,m+1}$ for doing this. The agent then purchases one m-maturity bond and shorts an $m+1$-maturity bond. The total cost of setting up this portfolio, which is short one contract, long one bond, and short one bond, is zero. At time m, the agent receives income 1 for the m-maturity bond, and he invests this at the money market rate R_m. At time $m+1$, this investment yields $1 + R_m$. The agent pays 1 to cover the short position in the $m+1$ maturity bond and uses the remaining R_m to pay off the contract. Thus, the agent has hedged the short position in the contract that promises to pay R_m at time $m+1$. The initial capital required to set up this hedge is $B_{n,m} - B_{n,m+1}$, which is thus the no-arbitrage price of the contract. □

According to Theorems 6.3.2 and 6.3.5, the forward price at time n of the contract that delivers R_m at time $m+1$ is $F_{n,m}$. (Replace m by $m+1$ in (6.3.2) and set $S_n = B_{n,m+1}F_{n,m}$.) This is another way to regard the concept of "locking in" an interest rate. Suppose at time n you short a forward contract on the interest rate (i.e., you agree to receive $F_{n,m}$ at time $m+1$ in exchange for a payment of R_m at that time). It costs nothing to enter this forward contract. Then you have locked in the interest rate $F_{n,m}$ at time m. Indeed, at time m you can invest 1 at the variable interest rate R_m. At time $m+1$, this investment yields $1+R_m$, but you have agreed to pay R_m in exchange for a payment of $F_{n,m}$, so the net amount you have at time $m+1$ after these settlements is $1+F_{n,m}$. You have effectively invested 1 between times m and $m+1$ at an interest rate of $F_{n,m}$.

Definition 6.3.6. *Let m be given with $1 \le m \le N$. An m-period interest rate swap is a contract that makes payments $S_1, \dots, S_m$ at times $1, \dots, m$, respectively, where*

$$S_n = K - R_{n-1}, \; n = 1, \dots, m.$$

The fixed rate K is constant. The m-period swap rate SR_m is the value of K that makes the time-zero no-arbitrage price of the interest rate swap equal to zero.

An agent with the long ("receive fixed") swap position receives a constant payment at each time n, a payment that one can regard as a fixed interest payment K on a principal amount of 1, and the agent makes a variable interest rate payment R_{n-1} on the same principal amount. If the agent already has a loan on which he is making fixed interest rate payments, the long swap position effectively converts this to a variable interest rate loan. A short swap position effectively converts a variable interest rate loan to a fixed interest rate loan.

Theorem 6.3.7. *The time-zero no-arbitrage price of the m-period interest rate swap in Definition 6.3.6 is*

$$\text{Swap}_m = \sum_{n=1}^{m} B_{0,n}(K - F_{0,n-1}) = K\sum_{n=1}^{m} B_{0,n} - (1 - B_{0,m}). \tag{6.3.4}$$

In particular, the m-period swap rate is

$$SR_m = \frac{\sum_{n=1}^m B_{0,n}F_{0,n-1}}{\sum_{n=1}^m B_{0,n}} = \frac{1 - B_{0,m}}{\sum_{n=1}^m B_{0,n}}. \tag{6.3.5}$$

PROOF: The time-zero no-arbitrage price of the payment K at time n is $KB_{0,n}$ and, according to Theorem 6.3.5, the time-zero no-arbitrage price of the payment R_{n-1} at time n is $B_{0,n}F_{0,n-1}$. Therefore, the time-zero price of S_n is $B_{0,n}(K - F_{0,n-1})$. Summing, we obtain the middle term in (6.3.4). But

$$\sum_{n=1}^{m} B_{0,n}F_{0,n-1} = \sum_{n=1}^{m} (B_{0,n-1} - B_{0,n}) = 1 - B_{0,m}.$$

This observation gives us the right-hand side of (6.3.4). Setting the swap price equal to zero and solving for K, we obtain (6.3.5). □

Formula (6.3.4) for the no-arbitrage price of an m-period swap is of course also consistent with risk-neutral pricing. More specifically, we have from (6.3.4), (6.3.3), (6.2.4), and (6.2.2) that

$$\begin{aligned}
\text{Swap}_m &= \sum_{n=1}^{m} B_{0,n}(K - F_{0,n-1}) \\
&= \sum_{n=1}^{m} \left[KB_{0,n} - (B_{0,n-1} - B_{0,n})\right] \\
&= \sum_{n=1}^{m} \left[K\widetilde{\mathbb{E}}D_n - (\widetilde{\mathbb{E}}D_{n-1} - \widetilde{\mathbb{E}}D_n)\right] \\
&= \sum_{n=1}^{m} \left[K\widetilde{\mathbb{E}}D_n - (\widetilde{\mathbb{E}}[D_n(1+R_{n-1})] - \widetilde{\mathbb{E}}D_n)\right] \\
&= \sum_{n=1}^{m} \left[K\widetilde{\mathbb{E}}D_n - \widetilde{\mathbb{E}}(D_nR_{n-1})\right] \\
&= \widetilde{\mathbb{E}}\sum_{n=1}^{m} D_n(K - R_{n-1}).
\end{aligned} \tag{6.3.6}$$

The last expression is the risk-neutral price of the swap.

Theorems 6.3.2, 6.3.5, and 6.3.7 determine prices of fixed income derivative securities by constructing hedging portfolios for short positions in the securities. Whenever such hedging portfolios can be constructed, risk-neutral pricing is justified, and (6.3.6) is an example of this. In the remainder of this section, we consider risk-neutral pricing of interest rate caps and floors, and although short position hedges for these instruments can usually be constructed, we do not work out the detailed assumptions that guarantee this is possible. Exercise 6.4 provides an example of a hedge construction.

Definition 6.3.8. *Let m be given, with $1 \le m \le N$. An m-period interest rate cap is a contract that makes payments $C_1, \dots, C_m$ at times $1, \dots, m$, respectively, where*

$$C_n = (R_{n-1} - K)^+, \; n = 1, \dots, m.$$

An m-period interest rate floor *is a contract that makes payments* $F_1, \dots, F_m$ *at times* $1, \dots, m$, *respectively, where*

$$F_n = \left(K - R_{n-1}\right)^+, \ n = 1, \dots, m.$$

A contract that makes the payment C_n *at only one time* n *is called an* interest rate caplet, *and a contract that makes the payment* F_n *at only one time* n *is called an* interest rate floorlet. *The* risk-neutral price of an m-period interest rate cap *is*

$$\text{Cap}_m = \widetilde{\mathbb{E}} \sum_{n=1}^{m} D_n (R_{n-1} - K)^+, \tag{6.3.7}$$

and the risk-neutral price of an m-period interest rate floor *is*

$$\text{Floor}_m = \widetilde{\mathbb{E}} \sum_{n=1}^{m} D_n (K - R_{n-1})^+. \tag{6.3.8}$$

If one is paying the variable interest rate R_{n-1} at each time n on a loan of 1, then owning an interest rate cap effectively caps the interest rate at K. Whenever the interest owed is more than K, the cap pays the difference. Similarly, if one is receiving variable interest R_{n-1} at each time n on an investment of 1, then owning an interest rate floor effectively provides a guaranteed interest rate of at least K. Whenever the interest received is less than K, the floor pays the difference.

Note that

$$K - R_{n-1} + (R_{n-1} - K)^+ = (K - R_{n-1})^+.$$

In other words, at each time, the payoff of a portfolio holding a swap and a cap is the same as the payoff of a floor. It follows that

$$\text{Swap}_m + \text{Cap}_m = \text{Floor}_m. \tag{6.3.9}$$

In particular, if K is set equal to the m-period swap rate, then the cap and the floor have the same initial price.

Example 6.3.9. We return to the example of Figure 6.2.1 but now assign values to the interest rate in the tree as shown in Figure 6.3.1.

In Table 6.1 we show $D_1 = \frac{1}{1+R_0}$, $D_2 = \frac{1}{(1+R_0)(1+R_1)}$, and $D_3 = \frac{1}{(1+R_0)(1+R_1)(1+R_2)}$. The right-hand column of this table records the probability $\widetilde{\mathbb{P}}\{A_{\omega_1\omega_2}\}$; i.e., the first entry in this column is the probability of HH on the first two tosses.

We can then compute the time-zero bond prices:

$$B_{0,1} = \widetilde{\mathbb{E}} D_1 = 1,$$

$$B_{0,2} = \widetilde{\mathbb{E}} D_2 = \frac{6}{7}\cdot\frac{1}{3} + \frac{6}{7}\cdot\frac{1}{6} + \frac{5}{7}\cdot\frac{1}{4} + \frac{5}{7}\cdot\frac{1}{4} = \frac{11}{14},$$

$$B_{0,3} = \widetilde{\mathbb{E}} D_3 = \frac{3}{7}\cdot\frac{1}{3} + \frac{6}{7}\cdot\frac{1}{6} + \frac{4}{7}\cdot\frac{1}{4} + \frac{4}{7}\cdot\frac{1}{4} = \frac{4}{7}.$$

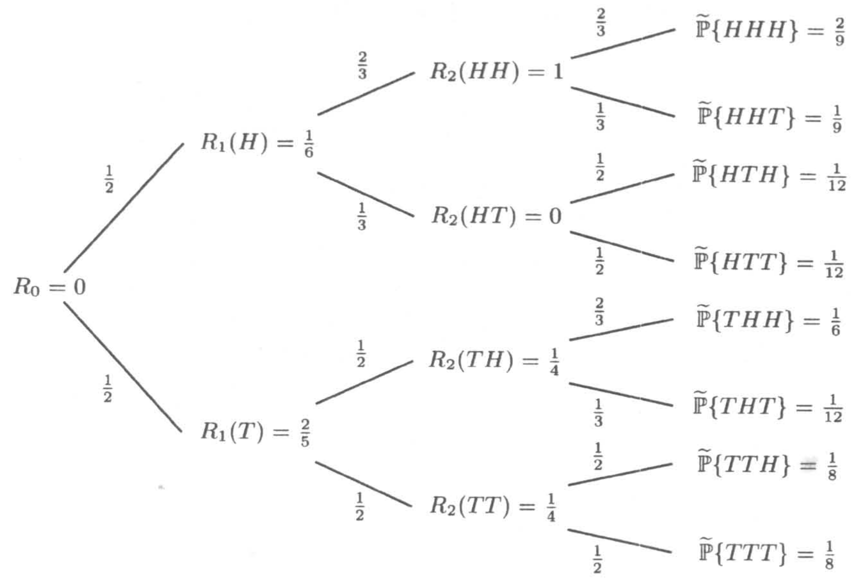

Fig. 6.3.1. A three-period interest rate model.

$\omega_1\omega_2$	$\frac{1}{1+R_0}$	$\frac{1}{1+R_1}$	$\frac{1}{1+R_2}$	D_1	D_2	D_3	$\widetilde{\mathbb{P}}$
HH	1	6/7	1/2	1	6/7	3/7	1/3
HT	1	6/7	1	1	6/7	6/7	1/6
TH	1	5/7	4/5	1	5/7	4/7	1/4
TT	1	5/7	4/5	1	5/7	4/7	1/4

Table 6.1.

The time-one bond prices are

$$B_{1,1} = 1,$$

$$B_{1,2}(H) = \frac{1}{D_1(H)} \widetilde{\mathbb{E}}_1[D_2](H) = \frac{6}{7}\cdot\frac{2}{3} + \frac{6}{7}\cdot\frac{1}{3} = \frac{6}{7},$$

$$B_{1,2}(T) = \frac{1}{D_1(T)} \widetilde{\mathbb{E}}_1[D_2](T) = \frac{5}{7}\cdot\frac{1}{2} + \frac{5}{7}\cdot\frac{1}{2} = \frac{5}{7},$$

$$B_{1,3}(H) = \frac{1}{D_1(H)} \widetilde{\mathbb{E}}_1[D_3](H) = \frac{3}{7}\cdot\frac{2}{3} + \frac{6}{7}\cdot\frac{1}{3} = \frac{4}{7},$$

$$B_{1,3}(T) = \frac{1}{D_1(T)} \widetilde{\mathbb{E}}_1[D_3](T) = \frac{4}{7}\cdot\frac{1}{2} + \frac{4}{7}\cdot\frac{1}{2} = \frac{4}{7}.$$

The time-two bond prices are

$$\begin{aligned}
B_{2,2} &= 1,\\
B_{2,3}(HH) &= \frac{1}{D_2(HH)}\widetilde{\mathbb{E}}_2[D_3](HH) = \frac{D_3(HH)}{D_2(HH)} = \frac{7}{6}\cdot\frac{3}{7} = \frac{1}{2},\\
B_{2,3}(HT) &= \frac{1}{D_2(HT)}\widetilde{\mathbb{E}}_2[D_3](HT) = \frac{D_3(HT)}{D_2(HT)} = \frac{7}{6}\cdot\frac{6}{7} = 1,\\
B_{2,3}(TH) &= \frac{1}{D_2(TH)}\widetilde{\mathbb{E}}_2[D_3](TH) = \frac{D_3(TH)}{D_2(TH)} = \frac{7}{5}\cdot\frac{4}{7} = \frac{4}{5},\\
B_{2,3}(TT) &= \frac{1}{D_2(TT)}\widetilde{\mathbb{E}}_2[D_3](TT) = \frac{D_3(TT)}{D_2(TT)} = \frac{7}{5}\cdot\frac{4}{7} = \frac{4}{5}.
\end{aligned}$$

Let us take $K = \frac{1}{3}$ and price the corresponding three-period interest rate cap. We record the payoff of this cap in the following table.

$\omega_1\omega_2$	R_0	$(R_0-\frac{1}{3})^+$	R_1	$(R_1-\frac{1}{3})^+$	R_2	$(R_2-\frac{1}{3})^+$
HH	0	0	1/6	0	1	2/3
HT	0	0	1/6	0	0	0
TH	0	0	2/5	1/15	1/4	0
HT	0	0	2/5	1/15	1/4	0

The price at time zero of the time-one, time-two, and time-three caplets, respectively, are

$$\begin{aligned}
\widetilde{\mathbb{E}}\Big[D_1\Big(R_0-\frac{1}{3}\Big)^+\Big] &= 0,\\
\widetilde{\mathbb{E}}\Big[D_2\Big(R_1-\frac{1}{3}\Big)^+\Big] &= \frac{5}{7}\cdot\frac{1}{15}\cdot\frac{1}{4}+\frac{5}{7}\cdot\frac{1}{15}\cdot\frac{1}{4} = \frac{1}{42},\\
\widetilde{\mathbb{E}}\Big[D_3\Big(R_2-\frac{1}{3}\Big)^+\Big] &= \frac{3}{7}\cdot\frac{2}{3}\cdot\frac{1}{3} = \frac{2}{21},
\end{aligned}\tag{6.3.10}$$

and so $\mathrm{Cap}_3 = 0 + \frac{1}{42} + \frac{2}{21} = \frac{5}{42}$.

6.4 Forward Measures

Let V_m be the payoff at time m of some contract (either a derivative security or some "primary" security such as a bond). The risk-neutral pricing formula for the price of this contract at times n prior to m involves the conditional expectation $\widetilde{\mathbb{E}}_n[D_mV_m]$. Indeed, the risk-neutral price of this security at time n is (see (6.2.8))

$$V_n = \frac{1}{D_n}\widetilde{\mathbb{E}}_n[D_mV_m],\ n = 0,1,\dots,m. \tag{6.4.1}$$

To compute the conditional expectation appearing in this formula when D_m is random, one would need to know the *joint* conditional distribution of D_m and V_m under the risk-neutral measure $\widetilde{\mathbb{P}}$. This makes the pricing of fixed income derivatives difficult.

One way around this dilemma is to build interest rate models using *forward measures* rather than the risk-neutral measure. The idea is to use the term D_m on the right-hand side of (6.4.1) as a Radon-Nikodým derivative to change to a different measure, and provided we then compute expectations under this different measure, the term D_m no longer appears. To make this idea precise, we begin with the following definition.

Definition 6.4.1. *Let m be fixed, with $1 \le m \le N$. We define*

$$Z_{m,m} = \frac{D_m}{B_{0,m}} \tag{6.4.2}$$

*and use $Z_{m,m}$ to define the m-*forward measure *$\widetilde{\mathbb{P}}^m$ by the formula*

$$\widetilde{\mathbb{P}}^m(\omega) = Z_{m,m}(\omega)\widetilde{\mathbb{P}}(\omega) \text{ for all } \omega\in\Omega.$$

To check that $\widetilde{\mathbb{P}}^m$ is really a probability measure, we must verify that it assigns probability one to Ω. This is the case because $\widetilde{\mathbb{E}}Z_{m,m} = 1$. Indeed,

$$\widetilde{\mathbb{P}}^m(\Omega) = \sum_{\omega\in\Omega}\widetilde{\mathbb{P}}^m(\omega) = \sum_{\omega\in\Omega}Z_{m,m}(\omega)\widetilde{\mathbb{P}}(\omega) = \widetilde{\mathbb{E}}Z_{m,m} = \frac{1}{B_{0,m}}\widetilde{\mathbb{E}}D_m = 1,$$

where the last equality follows from the definition of zero-coupon bond prices (see (6.2.3)). Following the development in Section 3.2, we may define the *Radon-Nikodým derivative process*

$$Z_{n,m} = \widetilde{\mathbb{E}}_nZ_{m,m},\ n = 0,1,\dots,m. \tag{6.4.3}$$

If V_m is a random variable depending on only the first m coin tosses, then according to Lemma 3.2.5,

$$\widetilde{\mathbb{E}}^mV_m = \widetilde{\mathbb{E}}[Z_{m,m}V_m]. \tag{6.4.4}$$

More generally, if $0\le n\le m$ and V_m depends on only the first m coin tosses, then according to Lemma 3.2.6,

$$\widetilde{\mathbb{E}}^m_n[V_m] = \frac{1}{Z_{n,m}}\widetilde{\mathbb{E}}_n[Z_{m,m}V_m]. \tag{6.4.5}$$

In the present context, with $Z_{m,m}$ defined by (6.4.2), equation (6.2.5) shows that equation (6.4.3) may be rewritten as

$$Z_{n,m} = \frac{D_nB_{n,m}}{B_{0,m}},\ n = 0,\dots,m. \tag{6.4.6}$$

Using (6.4.6) in (6.4.4) and (6.4.5), we obtain the following result.

Theorem 6.4.2. *Let m be fixed, with $1 \le m \le N$, and let $\widetilde{\mathbb{P}}^m$ denote the m-forward measure. If V_m is a random variable depending on only the first m coin tosses, then*

$$\widetilde{\mathbb{E}}^m[V_m] = \frac{1}{B_{0,m}}\widetilde{\mathbb{E}}[D_m V_m]. \tag{6.4.7}$$

More generally, if V_m depends on only the first m coin tosses, then

$$\widetilde{\mathbb{E}}_n^m[V_m] = \frac{1}{D_n B_{n,m}}\widetilde{\mathbb{E}}_n[D_m V_m], \quad n = 0, 1, \ldots, m. \tag{6.4.8}$$

Computation of the left-hand side of (6.4.8) does not require us to know the correlation between V_m and D_m. Rather, we only need to know the conditional distribution of V_m under the m-forward measure $\widetilde{\mathbb{P}}^m$. Thus, the left-hand side of (6.4.8) is often easier to compute than the term $\widetilde{\mathbb{E}}_n[D_m V_m]$ appearing on the right-hand side. Note, however, that the m-forward measure is useful for pricing only those securities that pay off at time m, not at other times.

From (6.4.8) and (6.4.1), we have

$$\widetilde{\mathbb{E}}_n^m[V_m] = \frac{V_n}{B_{n,m}}, \quad n = 0, 1, \ldots, m. \tag{6.4.9}$$

In other words, $\widetilde{\mathbb{E}}_n^m[V_m]$ is the price at time n of any derivative security or asset paying V_m at time m *denominated in units of the zero-coupon bond maturing at time m*. This is the *forward price* of the security or asset given in Theorem 6.3.2. The price of an asset denominated this way is a martingale under the forward measure $\widetilde{\mathbb{P}}^m$, as one can readily see by applying iterated conditioning to the left-hand side of (6.4.9). In conclusion, we see that *m-forward prices of (nondividend-paying) assets are martingales under the forward measure $\widetilde{\mathbb{P}}^m$.*

Example 6.4.3. Note from formula (6.4.2) that, like D_m, $Z_{m,m}$ depends on only the first $m-1$ coin tosses. We set $m = 3$ in Definition 6.4.1 and use the data in Example 6.3.9 so that $Z_{3,3}$ given by (6.4.2) is

$$Z_{3,3}(HH) = \frac{D_3(HH)}{B_{0,3}} = \frac{7}{4}\cdot\frac{3}{7} = \frac{3}{4}, \quad Z_{3,3}(HT) = \frac{D_3(HT)}{B_{0,3}} = \frac{7}{4}\cdot\frac{6}{7} = \frac{3}{2},$$
$$Z_{3,3}(TH) = \frac{D_3(TH)}{B_{0,3}} = \frac{7}{4}\cdot\frac{4}{7} = 1, \quad Z_{3,3}(TT) = \frac{D_3(TT)}{B_{0,3}} = \frac{7}{4}\cdot\frac{4}{7} = 1.$$

Note that $\widetilde{\mathbb{E}}Z_{3,3} = \frac{3}{4}\cdot\frac{1}{3} + \frac{3}{2}\cdot\frac{1}{6} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} = 1$, as it should. For each $\omega \in \Omega$, the values of $\widetilde{\mathbb{P}}(\omega)$, $Z_{3,3}(\omega)$, and $\widetilde{\mathbb{P}}^3(\omega) = Z_{3,3}(\omega)\widetilde{\mathbb{P}}(\omega)$ are given in the following table.

$\omega_1\omega_2\omega_3$	$\widetilde{\mathbb{P}}(\omega_1\omega_2\omega_3)$	$Z_{3,3}(\omega_1\omega_2\omega_3)$	$\widetilde{\mathbb{P}}^3(\omega_1\omega_2\omega_3)$
HHH	$\frac{2}{9}$	$\frac{3}{4}$	$\frac{1}{6}$
HHT	$\frac{1}{9}$	$\frac{3}{4}$	$\frac{1}{12}$
HTH	$\frac{1}{12}$	$\frac{3}{2}$	$\frac{1}{8}$
HTT	$\frac{1}{12}$	$\frac{3}{2}$	$\frac{1}{8}$
THH	$\frac{1}{6}$	1	$\frac{1}{6}$
THT	$\frac{1}{12}$	1	$\frac{1}{12}$
TTH	$\frac{1}{8}$	1	$\frac{1}{8}$
TTT	$\frac{1}{8}$	1	$\frac{1}{8}$

As in Example 6.2.1 and Figure 6.2.1, we can compute the $\widetilde{\mathbb{P}}^m$-conditional probabilities of getting H and T at each node in the tree representing the interest rate evolution. These conditional probabilities are shown in Figure 6.4.1.

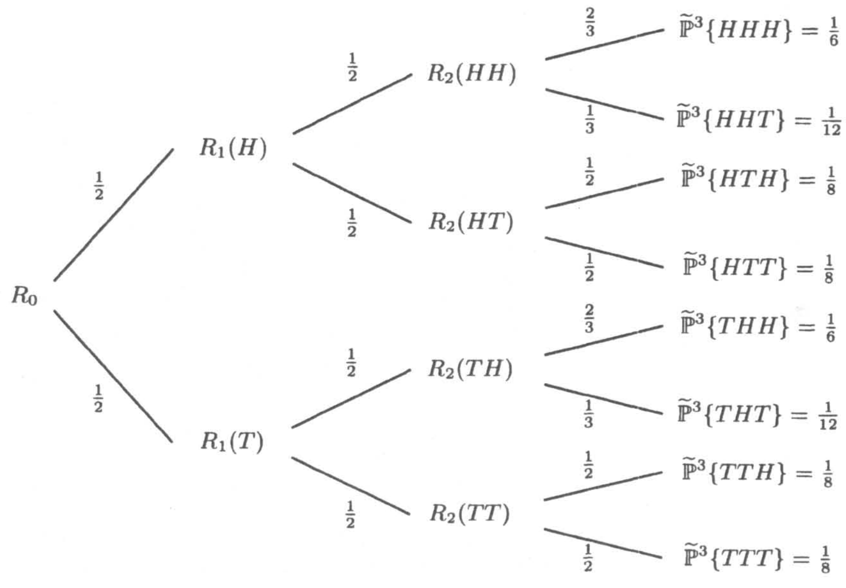

Fig. 6.4.1. Interest rate model and the $\widetilde{\mathbb{P}}^3$ transition probabilities.

We now compute the left-hand side of (6.4.8) with $V_3 = \left(R_2 - \frac{1}{3}\right)^+$, the payoff at time three of the caplet on the interest rate set at time two. In this case, V_3 depends on only the first two coin tosses and, in fact,

$$V_3(\omega_1\omega_2) = \begin{cases} \frac{2}{3}, & \text{if } \omega_1 = H, \omega_2 = H, \\ 0, & \text{otherwise.} \end{cases}$$

Since V_3 depends on only the first two tosses, we have $\widetilde{\mathbb{E}}_2^3[V_3] = V_3$. Using the probabilities shown in Figure 6.4.1, we compute

$$\begin{aligned} \widetilde{\mathbb{E}}_1^3[V_3](H) &= \frac{1}{2}V_3(HH) + \frac{1}{2}V_3(HT) = \frac{1}{3}, \\ \widetilde{\mathbb{E}}_1^3[V_3](T) &= \frac{1}{2}V_3(TH) + \frac{1}{2}V_3(TT) = 0, \\ \widetilde{\mathbb{E}}^3[V_3] &= \frac{1}{4}V_3(HH) + \frac{1}{4}V_3(HT) + \frac{1}{4}V_3(TH) + \frac{1}{4}V_3(TT) = \frac{1}{6}. \end{aligned}$$

The process $\widetilde{\mathbb{E}}_n^3[V_3]$, $n = 0, 1, 2, 3$, is displayed in Figure 6.4.2. We note that

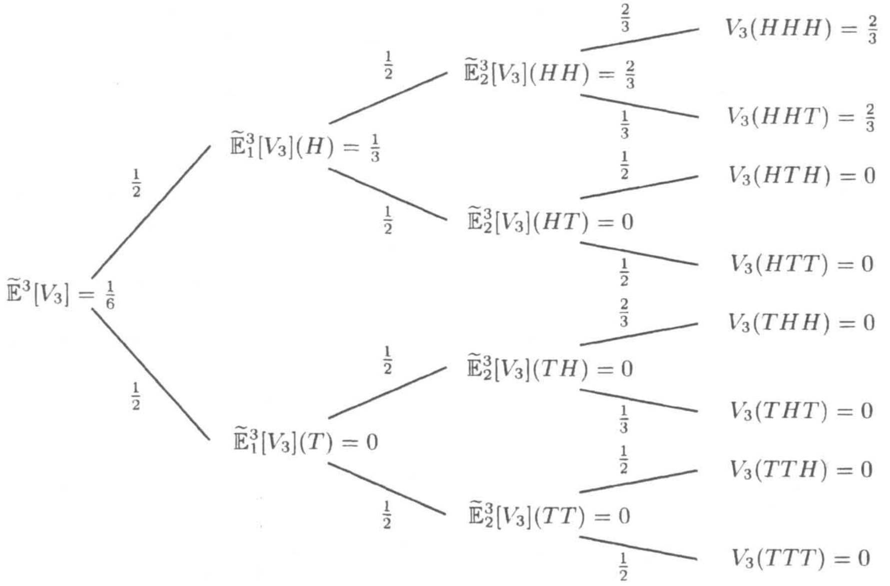

Fig. 6.4.2. The $\widetilde{\mathbb{P}}^3$ martingale $\widetilde{\mathbb{E}}_n^3[V_3]$.

this process is a martingale under $\widetilde{\mathbb{P}}^3$; the value at each node in the tree is the weighted average of the values at the two following nodes, using the $\widetilde{\mathbb{P}}^3$ transition probabilities shown on the links.

Finally, we use (6.4.9) to compute the time-zero price of the time-three caplet, which is

$$V_0 = B_{0,3}\widetilde{\mathbb{E}}^3[V_3] = \frac{4}{7} \cdot \frac{1}{6} = \frac{2}{21},$$

where we have used the bond price $B_{0,3} = \frac{4}{7}$ computed in Example 6.3.9. This agrees with (6.3.10). □

Example 6.4.4 (Ho-Lee model). In the Ho-Lee model [22], the interest rate at time n is

$$R_n(\omega_1 \dots \omega_n) = a_n + b_n \cdot \#H(\omega_1 \dots \omega_n),$$

where $a_0, a_1, \dots$ and $b_1, b_2, \dots$ are constants used to calibrate the model (i.e., make the prices generated by the model agree with market data). The risk-neutral probabilities are taken to be $\tilde{p} = \tilde{q} = \frac{1}{2}$.

In contrast to the kinds of numbers appearing in Examples 6.3.9 and 6.4.3, which were chosen to simplify the arithmetic, the Ho-Lee model can be used to generate numbers for practical applications. We repeat the computations of Examples 6.3.9 and 6.4.3 for the three-period Ho-Lee model shown in Figure 6.4.3 with $a_0 = 0.05$, $a_1 = 0.045$, $a_2 = 0.04$, and $b_1 = b_2 = 0.01$.

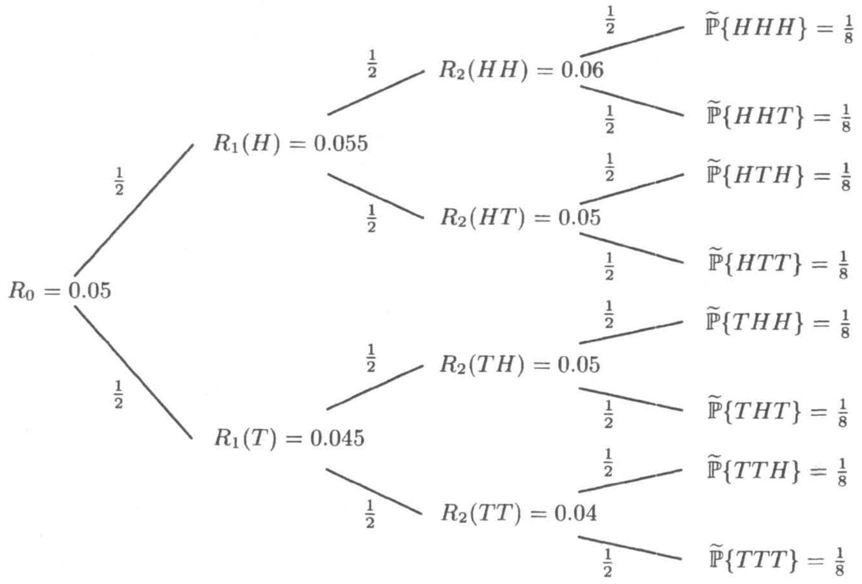

Fig. 6.4.3. A three-period Ho-Lee interest rate model.

In the following table, we show $D_1 = \frac{1}{1+R_0}$, $D_2 = \frac{1}{(1+R_0)(1+R_1)}$, and $D_3 = \frac{1}{(1+R_0)(1+R_1)(1+R_2)}$.

$\omega_1\omega_2$	$\frac{1}{1+R_0}$	$\frac{1}{1+R_1}$	$\frac{1}{1+R_2}$	D_1	D_2	D_3	$\widetilde{\mathbb{P}}$
HH	0.9524	0.9479	0.9434	0.9524	0.9027	0.8516	1/4
HT	0.9524	0.9479	0.9524	0.9524	0.9027	0.8597	1/4
TH	0.9524	0.9569	0.9524	0.9524	0.9114	0.8680	1/4
TT	0.9524	0.9569	0.9615	0.9524	0.9114	0.8763	1/4

We can then compute the time-zero bond prices:

$$B_{0,1} = \widetilde{\mathbb{E}}D_1 = 0.9524, \quad B_{0,2} = \widetilde{\mathbb{E}}D_2 = 0.9071, \quad B_{0,3} = \widetilde{\mathbb{E}}D_3 = 0.8639.$$

The time-one bond prices are $B_{1,1} = 1$ and

$$B_{1,2}(H) = \frac{1}{D_1(H)}\widetilde{\mathbb{E}}_1[D_2](H) = 0.9479,$$
$$B_{1,2}(T) = \frac{1}{D_1(T)}\widetilde{\mathbb{E}}_1[D_2](T) = 0.9569,$$
$$B_{1,3}(H) = \frac{1}{D_1(H)}\widetilde{\mathbb{E}}_1[D_3](H) = 0.8985,$$
$$B_{1,3}(T) = \frac{1}{D_1(T)}\widetilde{\mathbb{E}}_1[D_3](T) = 0.9158.$$

The time-two bond prices are $B_{2,2} = 1$ and

$$B_{2,3}(HH) = \frac{1}{D_2(HH)}\widetilde{\mathbb{E}}_2[D_3](HH) = 0.9434,$$
$$B_{2,3}(HT) = \frac{1}{D_2(HT)}\widetilde{\mathbb{E}}_2[D_3](HT) = 0.9524,$$
$$B_{2,3}(TH) = \frac{1}{D_2(TH)}\widetilde{\mathbb{E}}_2[D_3](TH) = 0.9524,$$
$$B_{2,3}(TT) = \frac{1}{D_2(TT)}\widetilde{\mathbb{E}}_2[D_3](TT) = 0.9615.$$

Let us take $K = 0.05$ and price the corresponding three-period interest rate cap. We record the payoff of this cap in the following table.

$\omega_1\omega_2$	R_0	$(R_0-0.05)^+$	R_1	$(R_1-0.05)^+$	R_2	$(R_2-0.05)^+$
HH	0.05	0	0.055	0.005	0.06	0.01
HT	0.05	0	0.055	0.005	0.05	0
TH	0.05	0	0.045	0	0.05	0
HT	0.05	0	0.045	0	0	0

The prices at time zero of the time-one, time-two, and time-three caplets, respectively, are

$$\widetilde{\mathbb{E}}\Big[D_1\Big(R_0 - 0.05\Big)^+\Big] = 0,$$
$$\widetilde{\mathbb{E}}\Big[D_2\Big(R_1 - 0.05\Big)^+\Big] = 0.002257,$$
$$\widetilde{\mathbb{E}}\Big[D_3\Big(R_2 - 0.05\Big)^+\Big] = 0.002129,$$

so the price of the cap consisting of these three caplets is $\mathrm{Cap}_3 = 0.004386$.

We next compute the 3-forward measure $\widetilde{\mathbb{P}}^3$. The Radon-Nikodým derivative of this measure with respect to the risk-neutral measure $\widetilde{\mathbb{P}}$ is

$$Z_{3,3}(HH) = \frac{D_3(HH)}{B_{0,3}} = 0.9858, \quad Z_{3,3}(HT) = \frac{D_3(HT)}{B_{0,3}} = 0.9952,$$
$$Z_{3,3}(TH) = \frac{D_3(TH)}{B_{0,3}} = 1.0047, \quad Z_{3,3}(TT) = \frac{D_3(TT)}{B_{0,3}} = 1.0144.$$

Note that $\widetilde{\mathbb{E}}Z_{3,3} = \frac{1}{4}\big(Z_{3,3}(HH)+Z_{3,3}(HT)+Z_{3,3}(TH)+Z_{3,3}(TT)\big) = 1$, as it should. For each $\omega \in \Omega$, the values of $\widetilde{\mathbb{P}}(\omega)$, $Z_{3,3}(\omega)$, and $\widetilde{\mathbb{P}}^3(\omega) = Z_{3,3}(\omega)\widetilde{\mathbb{P}}(\omega)$ are given in the following table.

$\omega_1\omega_2\omega_3$	$\widetilde{\mathbb{P}}(\omega_1\omega_2\omega_3)$	$Z_{3,3}(\omega_1\omega_2\omega_3)$	$\widetilde{\mathbb{P}}^3(\omega_1\omega_2\omega_3)$
HHH	$\frac{1}{8}$	0.9858	0.1232
HHT	$\frac{1}{8}$	0.9858	0.1232
HTH	$\frac{1}{8}$	0.9952	0.1244
HTT	$\frac{1}{8}$	0.9952	0.1244
THH	$\frac{1}{8}$	1.0047	0.1256
THT	$\frac{1}{8}$	1.0047	0.1256
TTH	$\frac{1}{8}$	1.0144	0.1268
TTT	$\frac{1}{4}$	1.0144	0.1268

The $\widetilde{\mathbb{P}}^3$ transition probabilities are shown in Figure 6.4.4.

We now compute the left-hand side of (6.4.8) with $V_3 = \big(R_2 - 0.05\big)^+$, the payoff at time three of the caplet on the interest rate set at time two. In this case, V_3 depends on only the first two coin tosses and, in fact,

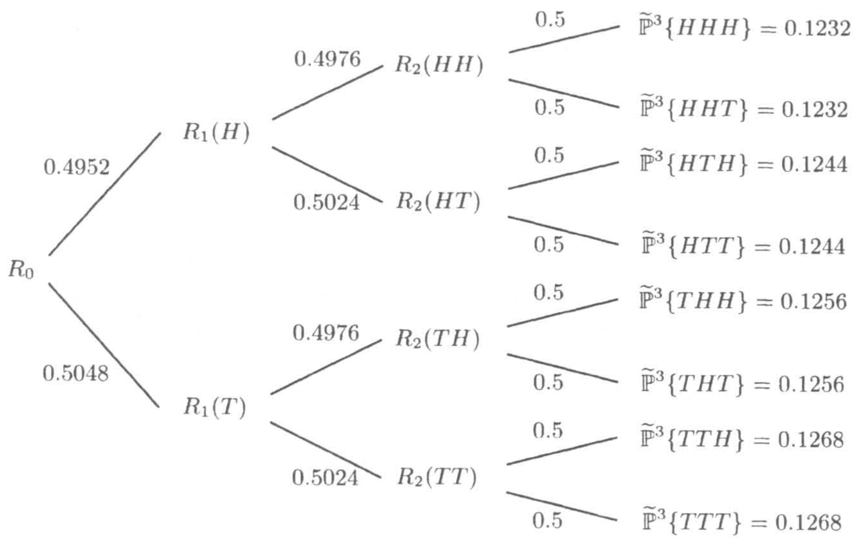

Fig. 6.4.4. The $\widetilde{\mathbb{P}}^3$ transition probabilities.

$$V_3(\omega_1\omega_2) = \begin{cases} 0.01, & \text{if } \omega_1 = H, \omega_2 = H, \\ 0, & \text{otherwise.} \end{cases}$$

Since V_3 depends on only the first two tosses, we have $\widetilde{\mathbb{E}}^3_2[V_3] = V_3$. Using the probabilities shown in Figure 6.4.4, we obtain the formulas shown in Figure 6.4.5. We note that this process is a martingale under $\widetilde{\mathbb{P}}^3$; the value at each node in the tree is the weighted average of the values at the two following nodes, using the $\widetilde{\mathbb{P}}^3$ transition probabilities shown on the links.

Finally, we use (6.4.9) to compute the time-zero price of the time-three caplet, which is $V_0 = B_{0,3}\widetilde{\mathbb{E}}^3[V_3] = (0.8639)(0.002464) = 0.002129$. □

6.5 Futures

Like a forward contract, a *futures contract* is designed to lock in a price for purchase or sale of an asset before the time of the purchase or sale. It is designed to address two shortcomings of forward contracts. The first of these is that on any date prior to a specified delivery date, there can be demand for forward contracts with that delivery date. Hence, efficient markets would for each delivery date need a multitude of forward contracts with different initiation dates. By contrast, the underlying process for futures contracts is a *futures price* tied to a delivery date but not an initiation date, and all agents, regardless of the date they wish to enter a futures contract, trade on

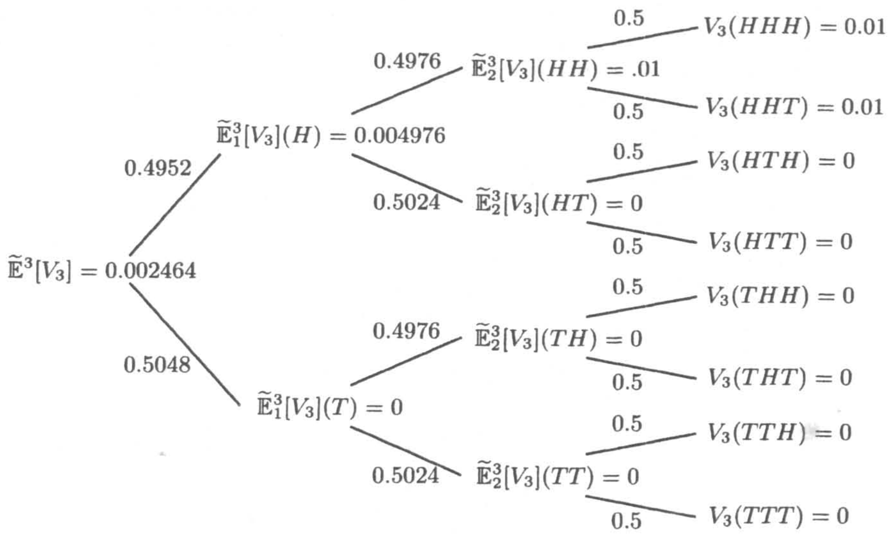

Fig. 6.4.5. The $\widetilde{\mathbb{P}}^3$ martingale $\widetilde{\mathbb{E}}^3_n[V_3]$.

the futures price for their desired delivery date. Secondly, although forward contracts have zero value at initiation, as the price of the underlying asset moves, the contract takes on a value that may be either positive or negative. In either case, one party may become concerned about default by the other party. Futures contracts, through *marking to market*, require agents to settle daily so that no agent ever has a large obligation to another.

Definition 6.5.1. *Consider an asset with price process $S_0, S_1, \ldots, S_N$ in the binomial interest rate model. For $0 \le m \le N$, the m-*futures price process *$\mathrm{Fut}_{n,m}$, $n = 0, 1, \ldots, m$, is an adapted process with the following properties:*

(i) $\mathrm{Fut}_{m,m} = S_m$;
(ii) For each n, $0 \le n \le m-1$, the risk-neutral value at time n of the contract that receives the payments $\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m}$ at time $k+1$ for all $k = n, \ldots, m-1$, is zero; i.e.,

$$\frac{1}{D_n}\widetilde{\mathbb{E}}_n\left[\sum_{k=n}^{m-1} D_{k+1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})\right] = 0. \tag{6.5.1}$$

An agent who at time n takes a long position in a futures contract with delivery at a later time m is agreeing to receive the payments $\mathrm{Fut}_{k+1,m} - F_{k,m}$ at times $k+1$, for $k = n, \ldots, m-1$ and then to take delivery of the asset at its market price S_m at time m. Any of the payments $\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m}$ can be negative, in which case the agent in the long position must pay money rather than receive it. The vehicle for these transfers of money is the margin account set up by the broker for the agent. At the end of each day, the broker either deposits money into or withdraws money from the margin account. If the account balance becomes too low, the broker issues a margin call, requiring the agent to either deposit more money in the account or close the futures position (see below).

An agent who holds a short futures position makes payments equal to $(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})$ and agrees to sell the asset at its market price S_m at time m. This agent likewise has a margin account.

The agent holding the long position between times n and m receives total payments

$$\sum_{k=n}^{m-1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m}) = \mathrm{Fut}_{m,m} - \mathrm{Fut}_{n,m} = S_m - \mathrm{Fut}_{n,m}\,.$$

After purchasing the asset at the market price S_m, the agent has a balance of

$$S_m - \mathrm{Fut}_{n,m} - S_m = -\mathrm{Fut}_{n,m}$$

from the futures trading and the asset purchase. In addition, the agent owns the asset. In effect, if one ignores the time value of money, the agent has paid $F_{n,m}$ and acquired the asset. In this sense, a price was locked in at time n for purchase at the later time m.

Condition (6.5.1) is designed so that at the time of initiation, the value of the futures contract is zero. Thus, it costs nothing to enter a long (or a short) futures position, apart from the requirement that a margin account be set up. Since (6.5.1) is required to hold for all n between 0 and $m-1$, the value of the futures contract is zero at all times. In particular, it costs nothing for an agent who holds a long position to "sell" his contract (i.e., to *close* the position). At that time, all payments (positive and negative) to the agent cease and the agent no longer is obligated to purchase the asset at the market price at time m. Similarly, an agent with a short futures position can close the position at any time at no cost.

Theorem 6.5.2. *Let m be given with $0 \le m \le N$. Then*

$$\mathrm{Fut}_{n,m} = \widetilde{\mathbb{E}}_n[S_m],\ n = 0, 1, \ldots, m \tag{6.5.2}$$

is the unique process satisfying the conditions of Definition 6.5.1.

PROOF: We first show that the right-hand side of (6.5.2) satisfies the conditions of Definition 6.5.1. It is clear that $\widetilde{\mathbb{E}}_m[S_m] = S_m$, which shows that the right-hand side of (6.5.2) satisfies Definition 6.5.1(i). To verify Definition 6.5.1(ii), it suffices to show that each term in the sum in (6.5.1) has $\widetilde{\mathbb{E}}_n$ conditional expectation zero; i.e.,

$$\widetilde{\mathbb{E}}_n\big[D_{k+1}(\widetilde{\mathbb{E}}_{k+1}[S_m] - \widetilde{\mathbb{E}}_k[S_m])\big] = 0,\ k = n, \ldots, m-1. \tag{6.5.3}$$

Because D_{k+1} depends on only the first k coin tosses, for $k = n, \ldots, m-1$, iterated conditioning implies

$$\begin{aligned}
\widetilde{\mathbb{E}}_n\Big[D_{k+1}\Big(\widetilde{\mathbb{E}}_{k+1}[S_m] - \widetilde{\mathbb{E}}_k[S_m]\Big)\Big] &= \widetilde{\mathbb{E}}_n\Big[\widetilde{\mathbb{E}}_k\Big[D_{k+1}\Big(\widetilde{\mathbb{E}}_{k+1}[S_m] - \widetilde{\mathbb{E}}_k[S_m]\Big)\Big]\Big]\\
&= \widetilde{\mathbb{E}}_n\Big[D_{k+1}\Big(\widetilde{\mathbb{E}}_k\big[\widetilde{\mathbb{E}}_{k+1}[S_m]\big] - \widetilde{\mathbb{E}}_k[S_m]\Big)\Big]\\
&= \widetilde{\mathbb{E}}_n\Big[\Big(D_{k+1}\big(\widetilde{\mathbb{E}}_k[S_m] - \widetilde{\mathbb{E}}_k[S_m]\big)\Big)\Big]\\
&= 0,
\end{aligned}$$

and (6.5.3) is established.

We next show that the right-hand side of (6.5.2) is the only process that satisfies the conditions of Definition 6.5.1. From (6.5.1), we have

$$\sum_{k=n}^{m-1}\widetilde{\mathbb{E}}_n\big[D_{k+1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})\big] = 0 \text{ for } n = 0, 1, \ldots, m-1. \tag{6.5.4}$$

For $n = 0, 1, \ldots, m-2$, we may replace n by $n+1$ in this equation and subtract to obtain

$$\begin{aligned}
0 = &\sum_{k=n}^{m-1}\widetilde{\mathbb{E}}_n\big[D_{k+1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})\big]\\
&- \sum_{k=n+1}^{m-1}\widetilde{\mathbb{E}}_{n+1}\big[D_{k+1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})\big].
\end{aligned} \tag{6.5.5}$$

If we now take conditional expectations $\widetilde{\mathbb{E}}_n$ in (6.5.5) and use iterated conditioning, we see that

$$\begin{aligned}
0 = &\sum_{k=n}^{m-1}\widetilde{\mathbb{E}}_n\big[D_{k+1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})\big]\\
&- \sum_{k=n+1}^{m-1}\widetilde{\mathbb{E}}_n\big[D_{k+1}(\mathrm{Fut}_{k+1,m} - \mathrm{Fut}_{k,m})\big]\\
= &\ \widetilde{\mathbb{E}}_n\big[D_{n+1}(\mathrm{Fut}_{n+1,m} - \mathrm{Fut}_{n,m})\big].
\end{aligned} \tag{6.5.6}$$

Setting $n = m-1$ in (6.5.4), we see that (6.5.6) also holds for $n = m-1$. Both $\mathrm{Fut}_{n,m}$ and D_{n+1} depend on only the first n coin tosses, and hence (6.5.6) reduces to

$$D_{n+1}\big(\widetilde{\mathbb{E}}_n[\mathrm{Fut}_{n+1,m}] - \mathrm{Fut}_{n,m}\big) = 0,$$

which yields

$$\widetilde{\mathbb{E}}_n[\mathrm{Fut}_{n+1,m}] = \mathrm{Fut}_{n,m},\ n = 0, 1, \ldots, m-1.$$

This is the martingale property under $\widetilde{\mathbb{P}}$ for $\mathrm{Fut}_{n,m}$, $n = 0, 1, \ldots, m$. But $\mathrm{Fut}_{m,m} = S_m$, and the martingale property implies (6.5.2). □

Corollary 6.5.3. *Let m be given with $0 \leq m \leq N$. Then $\mathrm{For}_{0,m} = \mathrm{Fut}_{0,m}$ if and only if D_m and S_m are uncorrelated under $\widetilde{\mathbb{P}}$. In particular, this is the case if the interest rate is not random.*

PROOF: From Theorem 6.3.2 we have

$$\mathrm{For}_{0,m} = \frac{S_0}{B_{0,m}} = \frac{S_0}{\widetilde{\mathbb{E}}D_m} = \frac{\widetilde{\mathbb{E}}[D_m S_m]}{\widetilde{\mathbb{E}}D_m}$$

and

$$\mathrm{Fut}_{0,m} = \widetilde{\mathbb{E}}S_m.$$

These two formulas agree if and only if $\widetilde{\mathbb{E}}[D_m S_m] = \widetilde{\mathbb{E}}D_m \cdot \widetilde{\mathbb{E}}S_m$, which is uncorrelatedness of D_m and S_m. □

Remark 6.5.4. We note in connection with Corollary 6.5.3 that if D_m and S_m are negatively correlated, so that $\widetilde{\mathbb{E}}[D_m S_m] < \widetilde{\mathbb{E}}D_m \cdot \widetilde{\mathbb{E}}S_m$, then $\mathrm{For}_{0,m} < \mathrm{Fut}_{0,m}$. This is the case when an increase in the asset price tends to be accompanied by an increase in the interest rate (so that there is a decrease in the discount factor). The long futures position benefits from this more than the long forward position since the long futures position receives an immediate payment to invest at the higher interest rate, whereas the forward contract has no settlement until the delivery date m. Because the long futures position is more favorable, the initial futures price is higher (i.e., the person using futures to lock in a purchase price locks in a higher purchase price than a person using a long forward position).

Example 6.5.5 (Black-Derman-Toy model). In the Black-Derman-Toy (BDT) model [4], the interest rate at time n is

$$R_n(\omega_1 \dots \omega_n) = a_n b_n^{\#H(\omega_1 \dots \omega_n)},$$

where the constants $a_0, a_1, \dots$ and $b_1, b_2 \dots$ are used to calibrate the model. The risk-neutral probabilities are taken to be $\tilde{p} = \tilde{q} = \frac{1}{2}$. With $a_n = \frac{0.05}{1.2^n}$ and $b_n = 1.44$, the three-period BDT model is shown in Figure 6.5.1. In this model, we have the following zero-coupon bond prices:

$$B_{0,2} = 0.9064, \quad B_{0,3} = 0.8620,$$
$$B_{1,2}(H) = 0.9434, \quad B_{1,2}(T) = 0.9600,$$
$$B_{1,3}(H) = 0.8893, \quad B_{1,3}(T) = 0.9210.$$

From these, we can compute the forward interest rates $F_{n,2} = \frac{B_{n,2} - B_{n,3}}{B_{n,3}}$, obtaining

$$F_{0,2} = 0.05147, \quad F_{1,2}(H) = 0.06089, \quad F_{1,2}(T) = 0.04231.$$

These are the forward prices for a contract that pays R_2 at time three. (The second subscript 2 on these forward interest rates denotes the time the rate

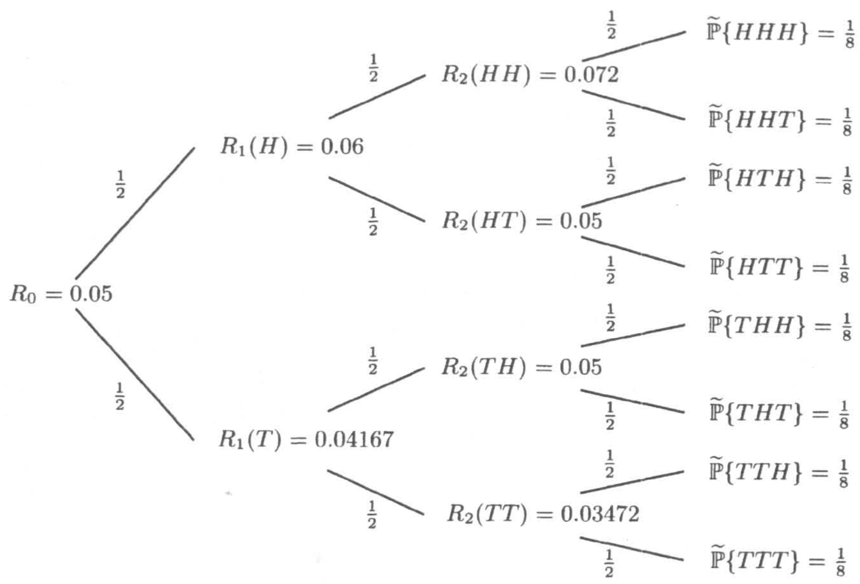

Fig. 6.5.1. A three-period Black-Derman-Toy interest rate model.

is set, not the time of payment.) The futures prices for the contract that pays R_2 at time three are given by $\mathrm{Fut}_{n,3} = \widetilde{\mathbb{E}}_n[R_2]$, and these turn out to be

$$\mathrm{Fut}_{0,3} = 0.05168, \quad \mathrm{Fut}_{1,3}(H) = 0.06100, \quad \mathrm{Fut}_{1,3}(T) = 0.04236.$$

The interest rate futures are slightly higher than the forward interest rates for the reason discussed in Remark 6.5.4.

6.6 Summary

Within the simple context of a binomial model, this chapter introduces several ideas that arise when interest rates are random. The prices of zero-coupon bonds are defined via the risk-neutral pricing formula (Definition 6.2.4), and it is shown that this method of pricing rules out arbitrage (Remark 6.2.7). Forward contracts, interest rate caps and floors, and futures contracts are then introduced. The last of these, appearing in Section 6.5, is related to forward contracts in Corollary 6.5.3.

Section 6.4 presents the concept of *forward measure*, an idea used in continuous-time models to simplify the pricing of interest rate caps and floors. We illustrate this idea in Example 6.4.3, but the full power of this approach is realized only in continuous-time models.

6.7 Notes

Many practical interest rate models are either formulated in discrete time or have discrete-time implementations (e.g., the models of Ho and Lee [22], Heath, Jarrow, and Morton [21], and Black, Derman, and Toy [4]). The concept of forward measures, which will be revisited in Volume II in a continuous-time setting, is due to Jamshidian [24] and Geman, El Karoui and Rochet [15]. These measures underlie the so-called *market models*, or *forward LIBOR models*, of interest rates developed by Sandmann and Sondermann [38] and Brace, Gatarek, and Musiela [6], models that are consistent with the *Black caplet formula*. The distinction between forward contracts and futures was pointed out by Margrabe [31] and Black [3]. No-arbitrage pricing of futures in a discrete-time model was developed by Cox, Ingersoll, and Ross [9] and Jarrow and Oldfield [25]. Exercise 6.7 is taken from Duffie [14], a highly recommended text.

6.8 Exercises

Exercise 6.1. Prove Theorem 2.3.2 when conditional expectation is defined by Definition 6.2.2.

Exercise 6.2. Verify that the discounted value of the static hedging portfolio constructed in the proof of Theorem 6.3.2 is a martingale under $\widetilde{\mathbb{P}}$.

Exercise 6.3. Let $0 \le n \le m \le N-1$ be given. According to the risk-neutral pricing formula, the contract that pays R_m at time $m+1$ has time-n price $\frac{1}{D_n}\widetilde{\mathbb{E}}_n[D_{m+1}R_m]$. Use the properties of conditional expectations to show that this gives the same result as Theorem 6.3.5, i.e.,

$$\frac{1}{D_n}\widetilde{\mathbb{E}}_n[D_{m+1}R_m] = B_{n,m} - B_{n,m+1}.$$

Exercise 6.4. Using the data in Example 6.3.9, this exercise constructs a hedge for a short position in the caplet paying $(R_2 - \frac{1}{3})^+$ at time three. We observe from the second table in Example 6.3.9 that the payoff at time three of this caplet is

$$V_3(HH) = \frac{2}{3}, \quad V_3(HT) = V_3(TH) = V_3(TT) = 0.$$

Since this payoff depends on only the first two coin tosses, the price of the caplet at time two can be determined by discounting:

$$V_2(HH) = \frac{1}{1+R_2(HH)}V_3(HH) = \frac{1}{3}, \quad V_2(HT) = V_2(TH) = V_2(TT) = 0.$$

Indeed, if one is hedging a short position in the caplet and has a portfolio valued at $\frac{1}{3}$ at time two in the event $\omega_1 = H, \omega_2 = H$, then one can simply invest this $\frac{1}{3}$ in the money market in order to have the $\frac{2}{3}$ required to pay off the caplet at time three.

In Example 6.3.9, the time-zero price of the caplet is determined to be $\frac{2}{21}$ (see (6.3.10)).

(i) Determine $V_1(H)$ and $V_1(T)$, the price at time one of the caplet in the events $\omega_1 = H$ and $\omega_1 = T$, respectively.
(ii) Show how to begin with $\frac{2}{21}$ at time zero and invest in the money market and the maturity two bond in order to have a portfolio value X_1 at time one that agrees with V_1, regardless of the outcome of the first coin toss. Why do we invest in the maturity two bond rather than the maturity three bond to do this?
(iii) Show how to take the portfolio value X_1 at time one and invest in the money market and the maturity three bond in order to have a portfolio value X_2 at time two that agrees with V_2, regardless of the outcome of the first two coin tosses. Why do we invest in the maturity three bond rather than the maturity two bond to do this?

Exercise 6.5. Let m be given with $0 \le m \le N-1$, and consider the forward interest rate

$$F_{n,m} = \frac{B_{n,m} - B_{n,m+1}}{B_{n,m+1}}, \quad n = 0, 1, \ldots, m.$$

(i) Use (6.4.8) and (6.2.5) to show that $F_{n,m}$, $n = 0, 1, \ldots, m$, is a martingale under the $(m+1)$-forward measure $\widetilde{\mathbb{P}}^{m+1}$.
(ii) Compute $F_{0,2}$, $F_{1,2}(H)$, and $F_{1,2}(T)$ in Example 6.4.4 and verify the martingale property

$$\widetilde{\mathbb{E}}^3[F_{1,2}] = F_{0,2}.$$

Exercise 6.6. Let S_m be the price at time m of an asset in a binomial interest rate model. For $n = 0, 1, \ldots, m$, the forward price is $\mathrm{For}_{n,m} = \frac{S_n}{B_{n,m}}$ and the futures price is $\mathrm{Fut}_{n,m} = \widetilde{\mathbb{E}}_n[S_n]$.

(i) Suppose that at each time n an agent takes a long forward position and sells this contract at time $n+1$. Show that this generates cash flow $S_{n+1} - \frac{S_n B_{n+1,m}}{B_{n,m}}$ at time $n+1$.
(ii) Show that if the interest rate is a constant r and at each time n an agent takes a long position of $(1+r)^{m-n-1}$ forward contracts, selling these contracts at time $n+1$, then the resulting cash flow is the same as the difference in the futures price $\mathrm{Fut}_{n+1,m} - \mathrm{Fut}_{n,m}$.

Exercise 6.7. Consider a binomial interest rate model in which the interest rate at time n depends on only the number of heads in the first n coin tosses. In other words, for each n there is a function $r_n(k)$ such that

$$R_n(\omega_1 \ldots \omega_n) = r_n(\#H(\omega_1 \ldots \omega_n)).$$

Assume the risk-neutral probabilities are $\tilde{p} = \tilde{q} = \frac{1}{2}$. The Ho-Lee model (Example 6.4.4) and the Black-Derman-Toy model (Example 6.5.5) satisfy these conditions.

Consider a derivative security that pays 1 at time n if and only if there are k heads in the first n tosses; i.e., the payoff is $V_n(k) = \mathbb{I}_{\{\#H(\omega_1 \ldots \omega_n)=k\}}$. Define $\psi_0(0) = 1$ and, for $n = 1, 2, \ldots$, define

$$\psi_n(k) = \widetilde{\mathbb{E}}[D_n V_n(k)], \quad k = 0, 1, \ldots, n,$$

to be the price of this security at time zero. Show that the functions $\psi_n(k)$ can be computed by the recursion

$$\psi_{n+1}(0) = \frac{\psi_n(0)}{2(1 + r_n(0))},$$

$$\psi_{n+1}(k) = \frac{\psi_n(k-1)}{2(1 + r_n(k-1))} + \frac{\psi_n(k)}{2(1 + r_n(k))}, \quad k = 1, \ldots, n,$$

$$\psi_{n+1}(n+1) = \frac{\psi_n(n)}{2(1 + r_n(n))}.$$

A

Proof of Fundamental Properties of Conditional Expectations

This appendix provides the proof of Theorem 2.3.2 of Chapter 2, which is restated below.

Theorem 2.3.2 (Fundamental properties of conditional expectations). *Let N be a positive integer, and let X and Y be random variables depending on the first N coin tosses. Let $0 \le n \le N$ be given. The following properties hold.*

(i) ***Linearity of conditional expectations.*** *For all constants c_1 and c_2, we have*

$$\mathbb{E}_n[c_1 X + c_2 Y] = c_1 \mathbb{E}_n[X] + c_2 \mathbb{E}_n[Y].$$

(ii) ***Taking out what is known.*** *If X actually depends only on the first n coin tosses, then*

$$\mathbb{E}_n[XY] = X \cdot \mathbb{E}_n[Y].$$

(iii) ***Iterated conditioning.*** *If $0 \le n \le m \le N$, then*

$$\mathbb{E}_n\Big[\mathbb{E}_m[X]\Big] = \mathbb{E}_n[X].$$

In particular, $\mathbb{E}\Big[\mathbb{E}_m[X]\Big] = \mathbb{E}X$.

(iv) ***Independence.*** *If X depends only on tosses $n+1$ through N, then*

$$\mathbb{E}_n[X] = \mathbb{E}X.$$

(v) ***Conditional Jensen's inequality.*** *If $\varphi(x)$ is a convex function of the dummy variable x, then*

$$\mathbb{E}_n[\varphi(X)] \ge \varphi(\mathbb{E}_n[X]).$$

PROOF: We start by recalling the definition of conditional expectation:

$$\mathbb{E}_n[X](\omega_1 \ldots \omega_n) = \sum_{\omega_{n+1} \ldots \omega_N} p^{\#H(\omega_{n+1} \ldots \omega_N)} q^{\#T(\omega_{n+1} \ldots \omega_N)} X(\omega_1 \ldots \omega_n \omega_{n+1} \ldots \omega_N).$$

Proof of (i):

$$
\begin{aligned}
&\mathbb{E}_n[c_1X + c_2Y](\omega_1\dots\omega_n)\\
&= \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}[c_1X(\omega_1\dots\omega_N) + c_2Y(\omega_1\dots\omega_N)]\\
&= c_1 \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}X(\omega_1\dots\omega_N)\\
&\quad +c_2 \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}Y(\omega_1\dots\omega_N)\\
&= c_1\mathbb{E}_n[X](\omega_1\dots\omega_n) + c_2\mathbb{E}_n[Y](\omega_1\dots\omega_n).
\end{aligned}
$$

Proof of (ii):

$$
\begin{aligned}
&\mathbb{E}_n[XY](\omega_1\dots\omega_n)\\
&= \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}X(\omega_1\dots\omega_n)Y(\omega_1\dots\omega_n\omega_{n+1}\dots\omega_N)\\
&=X(\omega_1\dots\omega_n) \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}Y(\omega_1\dots\omega_n\omega_{n+1}\dots\omega_N)\\
&= X(\omega_1\dots\omega_n)\mathbb{E}_n[Y](\omega_1\dots\omega_n).
\end{aligned}
$$

Proof of (iii): Denote $Z = E_m[X]$. Then Z actually depends on $\omega_1\omega_2\dots\omega_m$ only and

$$
\begin{aligned}
\mathbb{E}_n[\mathbb{E}_m[X]](\omega_1\dots\omega_n) &= \mathbb{E}_n[Z](\omega_1\dots\omega_n)\\
&= \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}Z(\omega_1\dots\omega_n\omega_{n+1}\dots\omega_m)\\
&= \sum_{\omega_{n+1}\dots\omega_m} p^{\#H(\omega_{n+1}\dots\omega_m)}q^{\#T(\omega_{n+1}\dots\omega_m)}Z(\omega_1\dots\omega_m)\\
&\quad\times \sum_{\omega_{m+1}\dots\omega_N} p^{\#H(\omega_{m+1}\dots\omega_N)}q^{\#T(\omega_{m+1}\dots\omega_N)}\\
&= \sum_{\omega_{n+1}\dots\omega_m} p^{\#H(\omega_{n+1}\dots\omega_m)}q^{\#T(\omega_{n+1}\dots\omega_m)}Z(\omega_1\dots\omega_m)\\
&= \sum_{\omega_{n+1}\dots\omega_m} p^{\#H(\omega_{n+1}\dots\omega_m)}q^{\#T(\omega_{n+1}\dots\omega_m)}\\
&\quad\times \sum_{\omega_{m+1}\dots\omega_N} p^{\#H(\omega_{m+1}\dots\omega_N)}q^{\#T(\omega_{m+1}\dots\omega_N)}X(\omega_1\dots\omega_N)\\
&= \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}X(\omega_1\dots\omega_N)\\
&= \mathbb{E}_n[X](\omega_1\dots\omega_n).
\end{aligned}
$$

Proof of (iv):

$$
\begin{aligned}
&\mathbb{E}_n[X](\omega_1\dots\omega_n)\\
&= \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}X(\omega_{n+1}\dots\omega_N)\\
&= \sum_{\omega_1\dots\omega_n} p^{\#H(\omega_1\dots\omega_n)}q^{\#T(\omega_1\dots\omega_n)}\\
&\qquad\cdot \sum_{\omega_{n+1}\dots\omega_N} p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}X(\omega_{n+1}\dots\omega_N)\\
&= \sum_{\omega_1\dots\omega_N} p^{\#H(\omega_1\dots\omega_N)}q^{\#T(\omega_1\dots\omega_N)}X(\omega_{n+1}\dots\omega_N)\\
&= \mathbb{E}X.
\end{aligned}
$$

Proof of (v): Let φ be a convex function, and denote by $\mathcal{L}$ the collection of all linear functions l that lie below φ (i.e., such that $l(y) \le \varphi(y)$ for all y). Then, as shown in the proof of Theorem 2.2.5,

$$\varphi(x) = \max_{l\in\mathcal{L}} l(x)$$

for all x.

For a random variable X and for all $\ell \in \mathcal{L}$, we have $\varphi(X) \ge \ell(X)$ and, consequently, $\mathbb{E}_n[\varphi(X)] \ge \mathbb{E}_n[\ell(X)]$. On the other hand, by property (i) (linearity), $\mathbb{E}_n[\ell(X)] = l(\mathbb{E}_n[X])$ so that $\mathbb{E}_n[\varphi(X)] \ge \ell(\mathbb{E}_n[X])$ for all $\ell \in \mathcal{L}$. Therefore

$$\mathbb{E}_n[\varphi(X)] \ge \max_{\ell\in\mathcal{L}} \ell(\mathbb{E}_n[X]) = \varphi(\mathbb{E}_n[X]).$$

References

1. ARROW, K. & DEBREU, G. (1954) Existence of equilibrium for a competitive economy, *Econometrica* **22**, 265–290.
2. BENSOUSSAN, A. (1984) On the theory of option pricing, *Acta Appl. Math.* **2**, 139–158.
3. BLACK, F. (1976) The pricing of commodity contracts, *J. Fin. Econ.* **3**, 167–179.
4. BLACK, F., DERMAN, E., & TOY, W. (1990) A one-factor model of interest rates and its application to treasury bond options, *Fin. Anal. J.* **46**, 33–39.
5. BLACK, F. & SCHOLES, M. (1973) The pricing of options and corporate liabilities, *J. Polit. Econ.* **81**, 637–659.
6. BRACE, A., GATAREK, D., & MUSIELA, M. (1997) The market model of interest-rate dynamics, *Math. Fin.* **7**, 127–154.
7. COX, J. C. & HUANG, C. (1989) Optimal consumption and portfolio policies when asset prices follow a diffusion process, *J. Econ. Theory* **49**, 33–83.
8. COX, J. C. & HUANG, C. (1991) A variational problem arising in financial economics, *J. Math. Econ.* **20**, 465–487.
9. COX, J. C., INGERSOLL, J. E., & ROSS, S. (1981) The relation between forward prices and futures prices, *J. Fin. Econ.* **9**, 321–346.
10. COX, J. C., INGERSOLL, J. E., & ROSS, S. (1985) A theory of the term structure of interest rates, *Econometrica* **53**, 385–407.
11. COX, J. C., ROSS, S., & RUBINSTEIN, M. (1979) Option pricing: a simplified approach, *J. Fin. Econ.* **3**, 145–166.
12. COX, J. C., ROSS, S., & RUBINSTEIN, M. (1985) *Options Markets*, Prentice-Hall, Englewood Cliffs, NJ.
13. DOOB, J. (1942) *Stochastic Processes*, J. Wiley & Sons, New York.
14. DUFFIE, D. (1992) *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ.
15. GEMAN, H., EL KAROUI, N., & ROCHET, J.-C. (1995) Changes of numéraire, changes of probability measure, and option pricing, *J. Appl. Prob.* **32**, 443–458.
16. HAKANSSON, N. (1970) Optimal investment and consumption strategies under risk for a class of utility functions, *Econometrica* **38**, 587–607.
17. HARRISON, J. M. & KREPS, D. M. (1979) Martingales and arbitrage in multi-period security markets, *J. Econ. Theory* **20**, 381–408.
18. HARRISON, J. M. & PLISKA, S. R. (1981) Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes Appl.* **11**, 215–260.

19. HEATH, D. (1995) A continuous-time version of Kulldorf's result, preprint, Department of Mathematical Sciences, Carnegie Mellon University.
20. HEATH, D., JARROW, R., & MORTON, A. (1992) Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation, *Econometrica* **60**, 77–105.
21. HEATH, D., JARROW, R., & MORTON, A. (1996) Bond pricing and the term structure of interest rates: a discrete time approximation, *Fin. Quant. Anal.* **25**, 419–440.
22. HO, T. & LEE, S. (1986) Term-structure movements and pricing interest rate contingent claims, *J. Fin.* **41**, 1011–1029.
23. HULL, J. & WHITE, A. (1990) Pricing interest rate derivative securities, *Rev. Fin. Stud.* **3**, 573–592.
24. JAMSHIDIAN, F. (1997) LIBOR and swap market models and measures, *Fin. Stochastics* **1**, 261–291.
25. JARROW, R. A. & OLDFIELD, G. S. (1981) Forward contracts and futures contracts, *J. Fin. Econ.* **9**, 373–382.
26. KARATZAS, I. (1988) On the pricing of American options, *Appl. Math. Optim.* **17**, 37–60.
27. KARATZAS, I., LEHOCZKY, J., & SHREVE, S. (1987) Optimal portfolio and consumption decisions for a "small investor" on a finite horizon, *SIAM J. Control Optim.* **25**, 1557–1586.
28. KARATZAS, I. & SHREVE, S. (1998) *Methods of Mathematical Finance*, Springer, New York.
29. KOLMOGOROV, A. N. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung, *Ergeb. Math.* **2**, No. 3. Reprinted by Chelsea Publishing Company, New York, 1946. English translation: *Foundations of Probability Theory*, Chelsea Publishing Co., New York, 1950.
30. KULLDORF, M. (1993) Optimal control of a favorable game with a time-limit, *SIAM J. Control Optim.* **31**, 52–69.
31. MARGRABE, W. (1978) A theory of forward and futures prices, preprint, Wharton School, University of Pennsylvania.
32. MERTON, R. (1969) Lifetime portfolio selection under uncertainty: the continuous-time case, *Rev. Econ. Statist.* **51**, 247–257.
33. MERTON, R. (1971) Optimum consumption and portfolio rules in a continuous-time model, *J. Econ. Theory* **3**, 373–413. Erratum: *ibid.* **6** (1973), 213–214.
34. MERTON, R. (1973) Theory of rational option pricing, *Bell J. Econ. Manage. Sci.* **4**, 141–183.
35. MERTON, R. (1973) An intertemporal capital asset pricing model, *Econometrica* **41**, 867–888.
36. MERTON, R. (1990) *Continuous-Time Finance*, Basil Blackwell, Oxford and Cambridge.
37. PLISKA, S. R. (1986) A stochastic calculus model of continuous trading: optimal portfolios, *Math. Oper. Res.* **11**, 371–382.
38. SANDMANN, K. & SONDERMANN, D. (1993) A term-structure model for pricing interest-rate derivatives, *Rev. Futures Markets* **12**, 392–423.
39. SHIRYAEV, A. N. (1978) *Optimal Stopping Rules*, Springer, New York.
40. SHIRYAEV, A. N. (1999) *Essentials of Stochastic Finance: Facts, Models, Theory*, World Scientific, Singapore.

41. SHIRYAEV, A. N., KABANOV, YU. M., KRAMKOV, D. O., & MELNIKOV, A. V. (1995) Towards the theory of pricing options of both European and American types. I. Discrete time, *Theory Prob. Appl.* **39**, 14–60.
42. VASICEK, O. (1977) An equilibrium characterization of the term structure, *J. Fin. Econ.* **5**, 177–188.
43. VILLE, J. (1939) *Étude Critique de la Notion du Collectif*, Gauthier-Villars, Paris.

Index